INFORMATION DESIGN IN SERVICE SYSTEMS AND ONLINE MARKETS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

David Alan Lingenbrink

May 2019

INFORMATION DESIGN IN SERVICE SYSTEMS AND ONLINE MARKETS

David Alan Lingenbrink, Ph.D.

Cornell University 2019

In mechanism design, the firm has an advantage over its customers in its knowledge of the state of the system, which can affect the utilities of all players. This poses the question: how can the firm utilize that information (and not additional financial incentives) to persuade customers to take actions that lead to higher revenue (or other firm utility)?

When the firm is constrained to "cheap talk," and cannot credibly commit to a manner of signaling, the firm cannot change customer behavior in a meaningful way. Instead, we allow firm to commit to how they will signal in advance. Customers can then trust the signals they receive and act on their realization. This thesis contains the work of three papers, each of which applies information design to service systems and online markets.

We begin by examining how a firm could signal a queue's length to arriving, impatient customers in a service system. We show that the choice of an optimal signaling mechanism can be written as a infinite linear program and then show an intuitive form for its optimal solution. We show that with the optimal fixed price and optimal signaling, a firm can generate the same revenue as it could with an observable queue and length-dependent variable prices.

Next, we study demand and inventory signaling in online markets: customers make strategic purchasing decisions, knowing the price will decrease if an item does not sell out. The firm aims to convince customers to buy now at a higher price. We show that the optimal signaling mechanism is public, and sends all customers the same information.

Finally, we consider customers whose ex ante utility is not simply their expected ex post utility, but instead a function of its distribution. We bound the number of signals needed for the firm to generate their optimal utility and provide a convex program reduction of the firm's problem.

BIOGRAPHICAL SKETCH

David was born on January 22, 1992 in Glendale, California and grew up there and in Sammamish, Washington. He attended Harvey Mudd College and completed a B.S. in Mathematics in May, 2010. There, he did a thesis under Michael Orrison and computational biology work under Ran Libeskind-Hadas. Three months after graduation, he began his Ph.D at Cornell.

ACKNOWLEDGEMENTS

First and foremost, I'd like to thank my parents, David and Lisa Lingenbrink. Without their guidance and continual support, I wouldn't be getting my doctorate or be half the person I am today. Through math flashcards and summer math programs, they were incredible at giving me avenues to explore without pushing me too far. I most thank them for being such good role models for me and providing a quality of person to shoot for.

Next, I'd like to thank my adviser, Krishnamurthy Iyer. He was incredibly patient with my growth as a researcher. When I compare my current abilities and writing to mine when I started, I realize how much work he did in cultivating that. The mentorship, guidance and support that Kris provided made me an unquestionably stronger mathematician and person.

I would also like to thank Siddhartha Banerjee and Robert Kleinberg for serving on my committee. They were always available for questions and provided important feedback on my work. The work of Chapter 4 was done in collaboration with Jerry Anunrojwong, as well, who is a brilliant researcher and attentive colleague.

From my earlier education, I'm very thankful to Michael Orrison, who gave me guidance throughout my studies and was very patient with me as I learned how to do self-directed research. Thanks, too, to Ran Libeskind-Hadas, who first introduced me to computer science and made my little academic discoveries feel important.

The last couple years of my work have been much smoother thanks to Margaret Moore, my incredibly supportive significant other. I thank her for all the ways she's made my life better.

I'm very grateful to the other ORIE Ph.D students, who have fostered a great community where I felt I could ask everybody a question about their (or my) work. Specifically, I'd like to thank:

- Andrew Daw, for being a friend through the strikes and gutters, ups and downs.
- Sam Gutekunst, for understanding being mentioned second and tolerating a whole lot.

- David Eckman, for being the funniest person in the office and for the many minor medical packs.
- Ben Grimmer, for pre-chewing and the Sam hot tub plan.
- Pamela Badian-Pessot, for being the heart of Peanut Office.
- Cory Girard, for being the spirit of Peanut Office.
- Calvin Wylie, for the year on Schuyler.
- Alberto Vera and Pu Yang, for all the ASoIaF theories.
- Ken Chong and Daniel Fleischman, for the puzzles and Only Connect.
- Chamsi Hssaine and Artur Gorokh, for the many notes on talks.
- Venus Lo, for all the baked goods.
- Matthew Zalesak, for all the other baked goods.
- All the other friends I've made in the department.
- The professors, who made grad school the template of a good workplace.
- The staff, for being so helpful and always making me feel welcome.

Additional thanks to my sister, Hannah Lingenbrink, who will always be my first friend. Thanks also to the members of my Cornell comedy group, The Skits. They made my Cornell experience much more fun and my adolescence much longer.

Chapter 2 is published as Lingenbrink and Iyer (2018a), and a preliminary version appeared at the 18th ACM Conference of Economics and Computation. A preliminary version (Lingenbrink and Iyer 2018b) of the work of Chapter 3 appeared at the 12th Workshop on the Economics of Networks, Systems and Computation. Chapter 4 was written in a collaboration with Jerry Anunrojwong (Anunrojwong et al. 2019).

I gratefully acknowledge the support of the NSF under the grant CMMI-1633920.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

> It takes time to persuade men to
> do even what is for their own good.
>
> — Thomas Jefferson

Information plays an important role in mechanism design. The level of information provided to users can dramatically change the actions they choose: with more information, users can make better informed choices. The mechanism designer has access to more information than the users, and can share it. Since the users' choices impact the utility of the mechanism designer, the mechanism designer strategically shares state information with users so the users choose actions that benefit the designer more frequently. By fully revealing the state information to each user, users will always take their best action. However, by instead combining some states (and telling users that the true state is within a certain set), the mechanism designer can strategically influence the behavior of users and increase their utility. The mechanism designer then faces a trade-off: an uninformative information sharing policy may never convince users to take an action, but one that is very informative may cause users to take less-preferred actions for some states. Information, then, has a value to users, but over-revealing state information may be sub-optimal in terms of designer's utility. The choice of the information provided to users in a mechanism is then very important, and careful information design can lead to significant increases in revenue and social welfare.

One prominent example of this trade-off is in the decisions of drivers in ride-sharing services such as Uber and Lyft. When a passenger makes a request to the service, they must indicate their destination, so the service can optimally match them with a driver. This information is not shared with the driver, who learns the destination upon picking up the passenger. This is done so that drivers cannot cherry-pick their fares and the service can guarantee a fast pick-up time for all passengers. This upsets many drivers, many of whom resort to practices banned by the service, such as calling the passenger and eliciting their destination or spoofing their own location, to attain the information. In this problem, destination information is valued by the drivers, but the ride-sharing service has found that sharing it would lead to lowered revenue from drivers strategically picking the best fares. However, ride-sharing services are experimenting with other solutions: in 2018, Uber started a pilot program to allow some drivers to occasionally see the expected duration and cardinal direction of a fare before accepting.

Our methods and results contribute to the emerging literature on Bayesian persuasion (Kamenica and Gentzkow 2011, Rayo and Segal 2010, Bergemann and Morris 2016a,b, 2018) that studies settings where an informed principal strategically chooses the amount of information to share with uninformed Bayesian agents to incentivize them to act in a desired manner. In contrast to the literature on *cheap talk* (Crawford and Sobel 1982), the distinctive feature in Bayesian persuasion is the assumption that the principal can *commit* to sharing information in a prespecified manner. The main insight is that, in general, the principal's optimal signal must obfuscate information by carefully coalescing favorable and unfavorable states of the agents. Kolotilin et al. (2016) extends this basic model to settings with privately informed agents who must report their types to the principal before receiving information. For a general methodological approach to Bayesian persuasion and information design in finite settings, see Bergemann and Morris (2018) and Taneva (2019).

In this thesis, we apply the techniques of Bayesian persuasion and information design to service systems and online markets. In Chapter 2, we consider a service system offering a service at a fixed price to impatient expected utility maximizing customers. Servers are limited, so arriving customers may choose to either wait in an unobservable queue to attain service or to leave the system forever. Before customers make their decision, the service provider can provide them with information on the current length of the queue[1]. We make no assumptions about the form this information may take; for example, valid forms of information are sending no information, sending the exact length, or sending a "busy"/"not busy" binary signal. Further, we make limited assumptions about the specifics of a customer's disutility from waiting; we only assume it's monotonically increasing and their net utility from the service becomes worse than their utility from leaving for especially slow service. Note that the customers' prior is dynamic; it is dependent on the choices of other customers. The problem for the service provider becomes the following: how should they signal to generate the most revenue from customers using the service? We first apply an argument analogous to the revelation principle in mechanism design: we show that to achieve the optimal revenue, the service provider need only send a binary signal recommendation to each customer, telling them what action to take. Next, we formulate the choice of an optimal signaling mechanism as the solution to an infinite linear program over the queue's steady state distribution. We then show that the optimal signaling mechanism must follow a simple form. When the customers' disutility from waiting is a linear function, we find an explicit form for the signaling mechanism. Finally, we show that for an optimal fixed price, the optimal signaling mechanism can generate the same revenue as the optimal length-dependent variable pricing mechanism where the queue is fully-observable.

In Chapter 3, we apply these methods to inventory and demand signaling in online retail. We consider a two period model where customers arrive in both periods to buy an item that is of limited stock. The price of the item is higher in the first period, but customers have a better chance of securing the item if they buy it early. The firm can observe both the total inventory available and the number of customers present at the first time period. The problem for the firm is: how can they signal the inventory and demand to all first-period customers to encourage them to buy early and maximize the firm's revenue? In contrast

[1]For simplicity, we let arrivals be according to a Poisson process and service times be exponential. Since these distributions are memoryless, the length of the queue is the *only* state information relevant to arriving customers.

to Chapter 2, the firm must now send a signal to many customers at once, and it is not immediately clear whether all customers should be sent the same information (we call this a *public* signaling mechanism). We notice that, like observed in the friendship paradox, customers believe the demand is stochastically higher than the firm believes it is. First, we again make a revelation principle style argument and restrict our attention to signaling mechanisms that recommend an action to every customer. We use this to, like in Chapter 2, write the choice of an optimal signaling mechanism as an infinite linear program. We establish that, when customers have homogeneous valuations, the optimal signaling mechanism is public. Additionally, we provide a polynomial time algorithm for computing it by posing it as an instance of the fractional knapsack problem. Note that, when customers value the item differently, the optimal signaling mechanism is not public, and we illustrate the gap between the optimal signaling mechanism and the optimal public signaling mechanism numerically in this chapter.

Finally, in Chapter 4, we consider customers whose utility cannot be written as an expectation over the utility of each possible state of the world. In particular, the customer's utility may be non-linear in their belief; such utility models arise, for example, when the receiver exhibits aversion to variability and risk in the payoff on choosing an action. In the presence of such non-linearity, the standard approach of using revelation-principle style arguments fails to provide an efficient characterization of the set of signals needed in the optimal signaling scheme. Our main contribution is to provide a theoretical framework, using results from convex analysis, to overcome this technical challenge. In particular, in general persuasion settings with risk-conscious agents, we prove that the sender's problem can be reduced to a convex optimization program. Furthermore, using this characterization, we obtain a bound on the number of signals needed in the optimal signaling scheme. We apply our methods to study a specific setting, namely binary persuasion, where the receiver has two possible action (0 and 1), and the sender always prefers the receiver taking action 1. Under a mild convexity assumption on the receiver's utility and using a geometric approach, we show that the convex program can be further reduced to a linear program. Furthermore, this linear program yields a canonical construction of the set of signals needed in an optimal signaling mechanism. In particular, this canonical set of signals only involves signals that fully reveal the state and signals that induce uncertainty between two states. We illustrate our results in the setting of signaling in a queueing system with customers whose utilities depend on the variance of their waiting times.

CHAPTER 2

OPTIMAL SIGNALING MECHANISMS IN UNOBSERVABLE QUEUES

All things come to those who wait.

Violet Fane

2.1 Introduction

In many services systems, where resources to serve users are often costly and limited, the user experience depends on the state of the system, namely the resource availability, wait times, the level of congestion, etc. As an example, in a call center, the wait-time until service affects a caller's experience. Similarly, in a ride-hailing service, the availability of drivers in a ride-requester's neighborhood directly influences the time until the requester begins her ride, thereby affecting her utility. When resource availability is too low or the wait times are high, the users in the system might prefer to not avail the service, and perhaps instead choose an outside option. For example, if the time until the beginning of a ride is too long, a ride-requester may instead choose to use public transport.

However, as compared to the service providers, the users of such services typically have far less information about the system state. A call center may know the number and the nature of other requests currently on hold, whereas such information is not available to a caller. Similarly, in a ride-hailing service, the platform has access to the number of drivers and their location around a ride-requester's neighborhood, whereas the requester *a priori* does not, unless informed by the platform. Without the current state information, a user may choose to obtain service when the system is in a poor state, and experience a low quality of service. One main goal of such systems is to minimize the occurrence of instances with poor quality of service, while maintaining revenue goals by providing service to a large number of users.

A common approach towards achieving this goal is to price the service based on the current system state. For example, in ride-hailing platforms, the price of obtaining rides often depends on the availability of the drivers in the platform. This price provides information to ride requesters, who may choose to not use the service if the price is too high. However, in some settings, practical considerations may render such state-dependent pricing infeasible or undesirable. This may be because there is no explicit price for the service being offered, or in other cases, the variability of prices may itself act as a source of user dissatisfaction.

When state-dependent pricing is infeasible or undesirable, a service provider may instead choose to share information about the system state directly to the users to help them decide whether or not to avail service. For example, a call-center may choose to make anticipated delay announcements to incoming callers to help them decide whether to stay on the line (Armony and Maglaras 2004a). Similarly, a ride-hailing service may choose to provide information about wait-times to help ride-requesters decide whether to hail rides. A natural question that arises then is how to effectively share information with users to reliably ensure they are satisfied with the service quality, while at the same time achieving revenue or profit goals. A secondary question is to quantify how the revenue so obtained compares with that under state-dependent pricing.

In this chapter, we study this problem of information sharing in the context of a service system offering service at a fixed price. Customers arriving at the system must decide whether to leave without obtaining service or to possibly join a queue to obtain service. The queue length is observable to the service provider but unobservable to the customers. Each customer is strategic and incurs a cost of waiting until service completion. Furthermore, the customers are Bayesian and incorporate any information shared by the service provider into their beliefs prior to making their decision. We consider a service provider interested in maximizing her expected revenue. We pose the following question in this setting: how should the service provider share information about the queue to incentivize participation, and maximize the expected revenue in the resulting customer equilibrium?

A central assumption in our model, as opposed to previous work (Allon et al. 2011), is that the service provider can *commit* to an information sharing mechanism. Without this commitment power, the service provider will always prefer to share (possibly false) information that maximizes the likelihood a customer joins the queue, and consequently, there cannot be any meaningful information transmission. On the other hand, as we show, by committing to a prespecified mechanism for information sharing, the service provider can credibly convey information about the state of the system.

Note that the set of all such mechanisms is quite complex. At one extreme, the service provider may choose to fully reveal the queue length to each arriving customers. At the other extreme, the service provider may choose not to disclose any information about the queue length to the customers. But, in between these two extremes, there exists a multitude of signaling mechanisms where the service provider sends a signal correlated with queue length to the customer. Moreover, each such choice of the signaling mechanism leads to the customers responding according to an equilibrium, and one must identify their equilibrium strategies in order to determine the resulting expected revenue. This task further exacerbates the complexity of identifying the optimal signaling mechanism.

The main contribution of this work is the rigorous formulation of the service providers' decision problem, and identifying, for general waiting costs, the structure of the optimal signaling mechanism. In particular, we show that the service provider's decision problem can be formulated as an infinite linear program, whose variables correspond to the steady-state distribution of the queue under a feasible signaling mechanism. By analyzing the linear program, we show that for any given fixed-price, there exists an optimal signaling mechanism that uses binary signals and has a threshold structure. This structure establishes that the optimal amount of information sharing requires the service provider to strategically provide ambiguous information about the queue, where the same signal is provided over a range of values of the queue length. In particular, the optimal signaling mechanism neither fully reveals nor fully conceals information about the system state.

We summarize our main results below.

(1) Linear programming formulation: We begin in Section 2.2 by formulating the service provider's decision problem as an optimization problem, where the customers' behavior is constrained to be in an equilibrium. In Section 2.3, we first use a revelation principle style argument (Fudenberg and Tirole 1991, Bergemann and Morris 2018) to show that it suffices to consider binary signaling mechanisms, where the signal the service provider sends is either "join" or "leave", and the customer equilibrium involves following the service provider's recommendation. Using this structural characterization of the set of signals, we show that the service provider's decision problem can be formulated as a linear program with a countable number of variables and constraints.

(2) Optimality of threshold mechanisms: Next, by analyzing this linear program, we establish in Section 2.4 that the optimal signaling mechanism has a threshold structure, where the service provider sends the "join" signal if the queue length is below some threshold, and "leave" otherwise. (In addition, at the threshold, the service provider may randomize.) We establish this result through a perturbative analysis, where any feasible solution is perturbed to a solution with better objective in two steps. Furthermore, in Section 2.6.1, for the special case of linear waiting costs, we use the structural characterization of the optimal mechanism to obtain closed-form expressions for the optimal value of the threshold for any fixed-price.

(3) Comparison of signaling with optimal state-dependent prices: Finally, in Section 2.5, we study the service provider's problem of setting the optimal fixed price in addition to subsequently choosing the optimal signaling mechanism. Interestingly, we find that with the

optimal choice of the fixed price and using the corresponding optimal signaling mechanism, the service provider can achieve the same revenue as with the optimal state-dependent pricing mechanism in an observable queue. (Hassin and Koshman (2017) obtain this result independently for the special case of linear waiting costs.) This suggests that in settings where state-dependent pricing is not feasible, the service provider can effectively use optimal signaling to achieve revenue comparable to those under state-dependent pricing.

This chapter provides a rigorous framework for analyzing the service provider's decision problem in a variety of related models that incorporate (exogenous) abandonments and customer heterogeneity, as we discuss in Sections 2.6.2 and 2.6.3. In particular, in these models, our framework leads to analogous linear programs whose solutions determine the optimal signaling mechanism. Our structural characterization of the optimal signaling mechanism continues to hold under abandonments. When customers are heterogeneous and their types are public, we show that the optimal mechanism may lack the threshold structure. However, we prove that the threshold structure of the optimal mechanism is restored if all customers types are charged the same price, or if the prices are set optimally.

2.1.1 Related work

Our work fits in the framework of Bayesian persuasion in a *dynamic* setting. Several recent papers fit this description. Kremer et al. (2014) study a setting where a group of agents must sequentially choose an action from a set of actions with unknown, but deterministic, rewards; a principal observes the reward obtained by each agent and may share information about this to the next agent in sequence, with the goal being to maximize the expected average reward across all agents. The central tension in this setting is that agents prefer to exploit given their information, whereas the principal seeks to balance exploration and exploitation. Papanastasiou et al. (2017) extend this model to allow for stochastic rewards in an infinite-horizon, decentralized multi-armed bandit setting with discounted rewards and characterize the optimal disclosure policy as a solution to a linear program. Mansour et al. (2016) study a similar model (and other more general settings) and propose a bandit algorithm that achieves asymptotically optimal regret in maximizing social welfare. Our work differs from these papers in two aspects. First, since these papers study learning in a bandit setting, the focus is on a transient analysis starting with exogenously specified priors. In contrast, we perform a steady-state analysis, which leads to the customers' prior beliefs arising endogenously in equilibrium. Second, these papers focus on social welfare maximization, whereas we analyze a setting where the principal seeks to maximize her own revenue. Finally, Ely (2017) studies Bayesian persuasion in a dynamic setting where a principal provides information about a stochastically evolving state to a myopic agent. In contrast to our work, the state evolution here is independent of the agent's actions.

Our work also ties into the long line of work on strategic behavior in queues in both observable and unobservable settings. In the seminar paper, Naor (1969) studies revenue and welfare maximizing through static pricing in an *observable* M/M/1 queue, where customers strategically choose to join or leave on arrival. Edelson and Hilderbrand (1975) study static pricing in an *unobservable* M/M/1 queue with strategic balking and observe that the revenue-maximizing static price equals welfare-maximizing static price. Chen and Frank (2001) study state-dependent pricing in an observable queue with homogeneous customers and prove that the revenue-optimal prices also maximize the social welfare. For more detailed discussions, see the book (Hassin and Haviv 2012), or the more recent extensive survey (Hassin 2016); in the following, we discuss few papers closely related to our model and results.

A number of papers have analyzed service systems where strategic customers are partially informed about system parameters and state (Burnetas and Economou 2007, Economou and Kanta 2008a,b), and the service provider makes announcements about delay and service quality. Armony and Maglaras (2004a) analyze a customer contact center where arriving customers choose among joining a queue to obtain service, leaving (never to return), and putting a service request for a call-back. Customers receive a state-dependent anticipated delay information before making their decision. (Armony and Maglaras (2004b) study a similar setting without the anticipated delay announcements.) The authors analyze a many-server, heavy-traffic regime and propose an asymptotically consistent delay announcement policy and an asymptotically optimal routing rule. Yu et al. (2017a) perform an empirical study on how delay announcements impact customer behavior using call-center data and observe that delay announcements directly affect customers' waiting costs. Cui and Veeraraghavan (2016) consider a setting where customers in an observable queue do not know the service parameters, such as the service rate, and have arbitrary beliefs about them. The authors compare the effects of revealing these parameters and find situations where the announcement of service parameters hurts consumer welfare. Pender et al. (2017, 2018) consider a setting where customers choosing between two queues are provided delayed queue-length information (or a moving-average of queue-lengths over a time window). They find that such information can lead to oscillations in the two queues if the delay is beyond a critical value. Hassin and Roet-Green (2017) study an unobservable queue where customers can obtain the queue-length information by paying a cost of inspection. The authors prove the existence and uniqueness of the equilibrium and study its properties for a range of inspection costs.

Our work is closely related to that of Allon et al. (2011), who consider an unobservable, single server queueing system where homogeneous customers with linear waiting costs choose to join or leave on arrival, after receiving a signal from the service provider. The authors assume that the service provider sends a deterministic signal at each queue state and focus on the setting of cheap talk, where the service provider cannot commit to the signaling mechanism. Essentially, the authors identify equilibria for the setting where the service provider and the customers choose their strategies simultaneously and study their properties for a range of settings differing in the alignment of the service provider's and the customers' incentives. Yu et al. (2017b) extend this model to include heterogeneous customers. In contrast to these papers, in our model, customers have general waiting costs, and the service provider can commit to the signaling mechanism. In other words, our model analyzes the Stackelberg setting where the service provider first selects (and commits to) a possibly randomized signaling mechanism, and the customers respond knowing the signaling mechanism. Finally, whereas they consider general objectives for the service provider, we focus on the setting where the service provider's goal is to maximize revenue.

Focusing on settings where the service provider has the power to commit, Hassin (1986) compares the social welfare between an observable queue and an unobservable queue, where customers have linear waiting costs and are charged revenue-maximizing static prices in each instance. The author notes that social welfare may be higher in the observable setting but not always. Guo and Zipkin (2007) study a similar setting where the service provider can commit to one of three specific signaling mechanisms (no information, total customers in the queue, or the exact total time needed to wait in the queue). Simhon et al. (2016) study a similar model under a specific class of signaling mechanisms, where the service provider reveals the queue length when it is below a threshold and reveals no information otherwise. They show that no such signaling mechanism can strictly increase the revenue over the full-information mechanism (in the overloaded regime) or the no-information mechanism (in the underloaded regime). In contrast, our model and methods do not *a priori* restrict the class of signaling mechanisms, and we show that typically the optimal signaling mechanism achieves strictly higher revenue than the full-information and the no-information mechanism. Furthermore, our analysis shows in fact that the service provider obtains higher revenue revealing the queue length when it is large and concealing it when it is short.

Recently, Hassin and Koshman (2017) analyze the case of linear waiting costs and observe that a threshold signaling mechanism, together with an optimal choice of the fixed price, achieves the optimal revenue. We obtain the same result for a broader class of customer waiting costs. Furthermore, we characterize the optimal mechanism for any exogenously specified fixed price.

Finally, in our results for linear waiting costs, the expression for the threshold in the optimal signaling mechanism involves the Lambert-W function. Borgs et al. (2014) obtain similar expressions involving the Lambert-W function for determining the optimal threshold in an admission control problem in observable queue.

2.2 Model

Our model consists of a service provider facing a sequence of potential customers who arrive according to a Poisson process with rate $\lambda > 0$. The service provider is capacity constrained, and consequently, customers seeking to obtain service are put into a queue. Each customer upon arrival must decide whether to join the queue to obtain service at a fixed price $p > 0$ or to leave without obtaining service. We focus on the setting where the queue is unobservable, i.e., the customers cannot directly see the state of the queue before deciding whether to join or leave. On the other hand, the service provider can observe the queue and may disclose information about its state to a customer upon her arrival.

We consider a setting where the queue is served by a single server. The service discipline in the queue is first-in-first-out (FIFO). Each customer's service requirement is distributed independently and identically as an exponential distribution and, without loss of generality, has unit mean. We restrict our attention to the setting where there is no abandonment: if the customer joins the queue upon arrival, they remain until service completion. We discuss different approaches to incorporate abandonment in our model in Section 2.6.2.

We make the assumption that the customers are homogeneous; we later discuss extensions to heterogeneous customers in Section 2.6.3. In particular, we represent the expected utility obtained by a customer upon joining the queue by a function $u(X)$, where X denotes the number of customers already in queue upon arrival of the customer. The net expected payoff obtained by the customer upon choosing to join the queue is then given by $h(X, p) \triangleq u(X) - p$. We normalize the payoff of leaving without obtaining service to zero.

We require that the function u is non-increasing in X, with $u(0) > 0$ and $\lim_{X\to\infty} u(X) < 0$. The first condition implies that customers incur a cost for waiting (longer) in queue. The latter two conditions are to avoid trivialities: the condition $u(0) > 0$ implies that a customer will prefer to join an empty queue if the price is low enough, whereas the final condition implies that for any $p \geq 0$, there exists an M such that $h(M, p) < 0$, making the customer prefer not joining the queue if she knows the queue length is larger than M. Given these assumptions, we restrict the values of p to the set $[0, u(0)]$, and for all $p \geq 0$, let M_p denote the smallest value of M for which we have $h(M, p) < 0$.

The arrival rate λ, the service requirement distribution, the customers' utility function $u(\cdot)$, and the fixed price p are common knowledge among the customers and the service provider.

2.2.1 Signaling mechanism

The service provider seeks to maximize her expected revenue and has two controls to achieve this goal: (1) the fixed price p at which the service is provided and (2) the information shared with each arriving customer regarding the state of the queue. To formally describe the latter, we next introduce the notion of a signaling mechanism. A signaling mechanism $\Sigma = (\mathcal{S}, \sigma)$ is composed of a set $\mathcal{S}$ of possible signals together with a mapping[1] $\sigma : \mathbb{N}_0 \times \mathcal{S} \to [0, 1]$, satisfying $\sum_{s\in\mathcal{S}} \sigma(n, s) = 1$ for each $n \in \mathbb{N}_0$. We interpret the mapping σ as follows: when a customer arrives to the system with X customers already in queue, the service provider sends a signal $s \in \mathcal{S}$ to the arriving customer with probability $\sigma(X, s)$.

[1] Here, and in the sequel, we let $\mathbb{N}_0$ denote the set of non-negative integers.

To illustrate our definition, we briefly discuss two natural signaling mechanisms that have been analyzed in the literature and serve as extreme benchmarks for comparison:

1. *No-information mechanism:* At one extreme, we have the no-information mechanism, where the service provider reveals no information about the queue state to arriving customers. This setting can be represented by a signaling mechanism $\Sigma = (\mathcal{S}, \sigma)$, where $\mathcal{S} = \{\emptyset\}$ and $\sigma(n, \emptyset) = 1$ for each $n \geq 0$. Edelson and Hilderbrand (1975) consider revenue maximization in unobservable queues (without any possibility of signaling).

2. *Fully revealing mechanism:* At the other extreme, we consider the fully-revealing mechanism, where the state of the queue is completely revealed to arriving customers. This setting can be represented in our model by a signal set $\mathcal{S} = \mathbb{N}_0$, and $\sigma(n, s) = 1$ if $s = n$ and 0 otherwise. The seminal paper by Naor (1969) studies the problem of revenue maximization in observable queues with strategic customers.

We assume that the service provider can commit to the signaling mechanism publicly, and that the signaling mechanism $\Sigma = (\mathcal{S}, \sigma)$ is common knowledge among the customers.

2.2.2 Customer equilibrium

The customers are strategic and Bayesian, and seek to maximize their total expected payoff given their beliefs. Given a signaling mechanism $\Sigma = (\mathcal{S}, \sigma)$, a pure strategy for a customer is a function $f : \mathcal{S} \to \mathcal{A} = \{0, 1\}$, that specifies, for each possible signal $s \in \mathcal{S}$, an action $f(s) \in \{0, 1\}$, where 1 denotes the action of joining the queue and 0 denotes the action of leaving without obtaining service. Similarly, a mixed strategy is specified by a function $f : \mathcal{S} \to [0, 1]$, where $f(s) \in [0, 1]$ denotes the probability that the customer will join the queue upon observing a signal $s \in \mathcal{S}$.

Recall that the service provider publicly commits to a signaling mechanism and seeks to maximize the expected revenue resulting from the customers' response. We model the customers' response as arising endogenously from an equilibrium. More precisely, we focus on the setting of a symmetric equilibrium where all customers follow the same (mixed) strategy. This is a mild assumption; in equilibrium, the customers' actions could possibly differ only at those signals under which they are indifferent between joining and leaving. To define the equilibrium notion, we consider a customer's decision problem when all other customers follow a given strategy.

Since customers are Bayesian, to describe a customer's decision problem, we must describe her beliefs. In particular, it is sufficient to describe the customer's *prior* belief about the state of the queue upon her arrival before receiving a signal from the service provider. (The customer's *posterior* belief after receiving a signal from the service provider is obtained via Bayes' rule.) Note that these prior beliefs are determined endogenously since the state of the queue upon a customer's arrival is dependent on the actions of all customers who arrived earlier. Since customer arrival is Poisson, using the PASTA property (Wolff 1982), we conclude that a customer upon arrival would see the queue in steady state. Consequently, in equilibrium, a customer's prior belief about the state of the queue must equal the queue's steady-state distribution.

Formally, given that all customers follow a strategy f and the service provider implements a signaling mechanism $\Sigma = (\mathcal{S}, \sigma)$, the queue evolves as a continuous time birth-death chain whose transition probabilities depend on f and σ. In particular, given there are n customers already in queue, a new customer enters the queue at rate $\lambda \sum_{s \in \mathcal{S}} \sigma(n, s) f(s)$, whereas a customer in service leaves the queue at rate 1. We restrict our attention to those customer strategies f for which the queue is stable[2]. Let $\pi_\infty(\Sigma, f)$ denote the steady state distribution of the queue under the signaling mechanism Σ and customers' strategies f. For notational

[2]We note that in any equilibrium (as defined below), the queue will be stable.

brevity, when the context is clear, we drop the explicit dependence on Σ and f to denote the steady state distribution by π_∞, and we let X_∞ denote a random variable distributed independently as π_∞.

Upon arrival, a customer's prior belief about the state of the queue is given by π_∞. Thus, after observing a signal s, the customer's expected payoff is given by $\mathbf{E}^{\Sigma,f}[h(X_\infty, p)|s]$, where $\mathbf{E}^{\Sigma,f}[\cdot|s]$ denotes expectation with respect to the customers' posterior beliefs conditional on the signaling mechanism Σ, the strategy f, and the observed signal s. From this expression, we conclude that the customer's optimal action is to join the queue if $\mathbf{E}^{\Sigma,f}[h(X_\infty, p)|s] > 0$, to leave if $\mathbf{E}^{\Sigma,f}[h(X_\infty, p)|s] < 0$, and any mixed action if $\mathbf{E}^{\Sigma,f}[h(X_\infty, p)|s] = 0$. This leads to the following definition of a customer equilibrium:

Definition 2.1. *Given a price p and a signaling mechanism Σ, a customer equilibrium is a strategy f satisfying for each $s \in \mathcal{S}$,*

$$f(s) = \begin{cases} 1 & \text{if } \mathbf{E}^{\Sigma,f}[h(X_\infty, p)|s] > 0; \\ 0 & \text{if } \mathbf{E}^{\Sigma,f}[h(X_\infty, p)|s] < 0, \end{cases} \tag{2.1}$$

and $f(s) \in [0,1]$ otherwise.

To illustrate, consider the setting where the customers' utility is linear $u(X) = 1 - c(X+1)$ for some $c \in (0,1)$ and let $p \in [0, 1-c]$. Under the fully-revealing mechanism, the equilibrium strategy is trivially given by $f(s) = 1$ if $s < \frac{1-c-p}{c}$ and 0 if $s > \frac{1-c-p}{c}$. (If $\frac{1-c-p}{c} \in \mathbb{N}_0$, then $f((1-c-p)/c)$ can take any value between 0 and 1.) On the other hand, under the no-information mechanism, the customer equilibrium strategy f can be computed to be $f(\emptyset) = \min\{\frac{1}{\lambda}(1 - \frac{c}{1-p}), 1\}$. (See Appendix A.2 for the details.)

2.2.3 Service provider's decision problem

Having defined the customer equilibrium, we are now ready to formally specify the service provider's decision problem. For a choice of the fixed price p and the signaling mechanism Σ, consider a customer equilibrium f. In steady state, the queue throughput is given by

$$\mathsf{Th}(\Sigma, f) \triangleq \mathbf{E}^{\Sigma,f}\left[\lambda \sum_{s\in\mathcal{S}} \sigma(X_\infty, s) f(s)\right] = \lambda \sum_{n=0}^{\infty} \pi_\infty(n) \sum_{s\in\mathcal{S}} \sigma(n,s) f(s).$$

The preceding equation follows from the fact that customers are arriving according to a Poisson process with rate λ and, upon arrival, see the queue in steady state. In steady state, the number of customers already in queue is n with probability $\pi_\infty(n)$, in which case the service provider sends a signal $s \in \mathcal{S}$ with probability $\sigma(n,s)$ and the customer joins the queue with probability $f(s)$. (Note that although the throughput $\mathsf{Th}(\Sigma, f)$ does not depend on the price p explicitly, there is an implicit dependence on p through the customer equilibrium f.) Thus, the service provider's expected revenue in equilibrium is given by

$$R(p, \Sigma, f) \triangleq p \cdot \mathsf{Th}(\Sigma, f).$$

The service provider's decision problem is then to choose a fixed price p and a signaling mechanism Σ in order to maximize her expected revenue in the resulting customer equilibrium[3] f:

$$\max_p \max_\Sigma R(p, \Sigma, f) \text{ subject to } f \text{ satisfying (2.1).} \tag{2.2}$$

Our main goal is to determine the optimal fixed price and to characterize the optimal signaling mechanism (if they exist) for the decision problem (2.2). As a first step in our analysis, we begin by studying the inner maximization problem, where the service provider seeks to choose an optimal signaling mechanism for a given (exogenously specified) fixed price p. For a given

[3]Note that in this formulation, we have ignored the possibility of the existence of multiple equilibria under a signaling mechanism. The right formulation would require that the service provider chooses, in addition to the signaling mechanism, a focal equilibrium f among all possible equilibria. Our results continue to hold under this formulation; we suppress the technical details for brevity and readability.

price, the service provider's problem can be equivalently cast as as throughput maximization problem, as specified below:

$$\max_{\Sigma} \mathsf{Th}(\Sigma, f) \text{ subject to } f \text{ satisfying (2.1).} \tag{2.3}$$

Subsequently, in Section 2.5, we address the problem of determining the optimal fixed price p.

2.3 Characterization of the signal space

There are two main difficulties in analyzing the decision problem (2.3). First, the space of possible signaling mechanisms is quite large. In particular, we have imposed no restrictions on the set $\mathcal{S}$. To make any progress, we must obtain some characterization of the set of possible signals that an optimal signaling mechanism might use. Second, given a particular signaling mechanism, one must characterize the customer equilibrium f^{σ}. This involves solving for a fixed point of an operator implicitly defined by (2.1), and for a general signaling mechanism, this could be a difficult problem. Hence, we first address these difficulties.

2.3.1 Equilibrium characterization

Towards the goal of characterizing the set of signals in an optimal signaling mechanism, we start by defining the notion of *equivalence* between two mechanisms and the respective customer equilibria.

Definition 2.2. *We say two signaling mechanisms $\Sigma_i = (\mathcal{S}_i, \sigma_i)$ and corresponding customer equilibria f_i, for $i = 1, 2$ are* equivalent *if they induce the same steady-state distribution, i.e., if $\pi_\infty(\Sigma_1, f_1) = \pi_\infty(\Sigma_2, f_2)$.*

We have the following lemma that states that it suffices to consider signaling mechanisms where the resulting customer equilibrium is pure. We provide the proof in Appendix A.1.

Lemma 2.1. *For any fixed price p, given a signaling mechanism $\Sigma = (\mathcal{S}, \sigma)$ and a customer equilibrium f, there exists a signaling mechanism $\Sigma_1 = (\mathcal{S}_1, \sigma_1)$ with customer equilibrium f_1 such that (1) (Σ_1, f_1) is equivalent to (Σ, f), and (2) f_1 is a pure strategy.*

Using the preceding lemma, we can further restrict the class of signaling mechanisms and customer equilibria to consider. We have the following lemma that states that it is enough for the service provider to consider mechanisms with binary signals with a specific customer equilibrium. The proof of the lemma uses a revelation-principle style argument (Fudenberg and Tirole 1991, Bergemann and Morris 2018); we include the proof in Appendix A.1 for the sake of completeness.

Lemma 2.2. *For any fixed price p, given a signaling mechanism $\Sigma = (\mathcal{S}, \sigma)$ and a customer equilibrium f, there exists an equivalent signaling mechanism $\Sigma_1 = (\mathcal{S}_1, \sigma_1)$ and customer equilibrium f_1, where $\mathcal{S}_1 = \{0, 1\}$ and $f_1(s) = s$ for $s \in \mathcal{S}_1$.*

Summing up the preceding two lemmas, we conclude that in order to determine an optimal signaling mechanism, it is sufficient to consider signaling mechanisms $\Sigma = (\mathcal{S}, \sigma)$ where $\mathcal{S} = \{0, 1\}$, and for which, the customer equilibrium is given by $f(s) = s$ for $s \in \{0, 1\}$. In other words, in the optimal signaling mechanism, the service provider sends a binary signal ("join" or "leave") depending on the queue length, and in the resulting equilibrium, each customer finds it optimal to follow the recommendation. We refer to this customer strategy as the *obedient* strategy (Bergemann and Morris 2016a, 2018) and the resulting equilibrium to be the obedient equilibrium.

Given this reduction, the service provider's decision problem, for any fixed price p, simplifies to identifying a mapping $\sigma : \mathbb{N}_0 \times \{0, 1\} \to [0, 1]$ (with the restriction that $\sigma(n, 0) = 1 - \sigma(n, 1)$

for each $n \in \mathbb{N}_0$) that maximizes the throughput:

$$\begin{aligned} \max_{\sigma} \quad & \mathbf{E}^{\sigma}[\lambda\sigma(X_\infty, 1)] \\ \text{subject to,} \quad & \mathbf{E}^{\sigma}[h(X_\infty, p) | s = 1] \geq 0, \\ & \mathbf{E}^{\sigma}[h(X_\infty, p) | s = 0] \leq 0. \end{aligned} \tag{2.4}$$

Here, the two inequalities impose the requirement that the customers find obedience to be optimal: when the signal $s = i$ is revealed to a customer, choosing action i is indeed an optimal action for her. We thus refer to the two constraints as *obedience constraints*. Note that, since we focus on the obedient equilibrium for the customers, and the signal space is fixed to be $\mathcal{S} = \{0, 1\}$, we simplify the notation and denote the expectation by $\mathbf{E}^{\sigma}$.

2.3.2 LP formulation

Observe that the preceding optimization problem (2.4) is quite complex: in addition to having an infinite number of variables $\{\sigma(n, i) : n \geq 0, i = 0, 1\}$, the constraints are highly non-linear. This non-linearity implies that optimizing directly would be difficult. In this section, we provide a reformulation of (2.4) as a linear program. This reformulation paves the way for analyzing the service provider's decision problem and for characterizing the structure of the optimal mechanism.

The main insight behind the reformulation is that instead of optimizing over the signaling mechanism σ, one can optimize directly over the resulting steady state distribution π_∞^{σ}. By doing so, the preceding non-linear optimization problem simplifies to the following linear program in $\{\pi^\sigma_\infty(n) : n \geq 0\}$, albeit with a countable number of variables and constraints:

$$
\begin{aligned}
\max_{\pi} \quad & \sum_{n=1}^{\infty} \pi_n \\
\text{subject to,} \quad & \sum_{n=1}^{\infty} \pi_n h(n-1,p) \geq 0 && \text{(2.5a)} \\
& \sum_{n=0}^{\infty} h(n,p) \left(\lambda \pi_n - \pi_{n+1}\right) \leq 0 && \text{(2.5b)} \\
& \lambda \pi_n - \pi_{n+1} \geq 0, \quad \text{for all } n \geq 0 && \text{(2.5c)} \\
& \sum_{i=0}^{\infty} \pi_i = 1, \qquad \pi_n \geq 0, \quad \text{for all } n \geq 0.
\end{aligned}
$$

To obtain this linear program, we first write the expectations in the obedience constraints of (2.4) as linear functions of the steady state distribution. The constraints (2.5c) are obtained from the detailed balance conditions $\pi^\sigma_\infty(n)\lambda\sigma(n,1) = \pi^\sigma_\infty(n+1)$ and using the fact that $\sigma(n,1) \in [0,1]$ for each $n \in \mathbb{N}_0$. We have the following lemma that relates the two optimization problems:

Lemma 2.3. *For every signaling mechanism $\sigma : \mathbb{N}_0 \times \{0,1\} \to [0,1]$ feasible for* (2.4)*, there exists a feasible solution $\{\pi_n : n \geq 0\}$ to* (2.5) *with the same objective value. Conversely, let $\{\pi_n : n \geq 0\}$ be feasible for* (2.5)*. Then the signaling mechanism $\sigma : \mathbb{N}_0 \times \{0,1\} \to [0,1]$, defined as $\sigma(n,1) = \frac{\pi_{n+1}}{\lambda \pi_n}$ if $\pi_n > 0$ and $\sigma(n,1) = 0$ otherwise, is feasible for* (2.4) *and has the same objective value.*

The preceding lemma not only allows us to optimize over the steady state distribution $\{\pi_n : n \geq 0\}$, but also provides a rule to determine $\sigma(n,1)$ from the optimal solution and hence recover the signaling mechanism. The proof of this lemma is given in Appendix A.1. With this reformulation of the service provider's problem, we are now ready to identify an optimal mechanism.

2.4 Structure of Optimal Mechanism

Note that the signaling mechanism still has to determine which binary signal to send at each queue length. In the following, we show that this problem has a simple structure. Towards that end, we introduce below the class of *threshold* mechanisms.

Definition 2.3. *We define a threshold mechanism* $\sigma^x : \mathbb{N}_0 \times \{0, 1\} \to [0, 1]$ *for* $x \in \mathbb{R}_+ \cup \{\infty\}$ *as follows: For* $x \in \mathbb{R}_+$ *we have*

$$\sigma^x(n, 1) \triangleq \begin{cases} 1 & \text{if } n < \lfloor x \rfloor; \\ x - \lfloor x \rfloor & \text{if } n = \lfloor x \rfloor; \\ 0 & \text{otherwise,} \end{cases}$$

Also, we define $\sigma^\infty(n, 1) = 1$ *for all* $n \geq 0$.

With this definition in place, we have our first main result:

Theorem 2.1. *For any fixed price* p*, there exists a threshold mechanism* σ^x *with* $x \in \mathbb{R}_+ \cup \{\infty\}$ *and* $x \geq M_p$ *that achieves the optimal revenue.*

Proof. The proof of the theorem involves three steps. First, analyzing the constraints of the linear program (2.5), we show that the optimal signaling mechanism would signal a customer to join the queue if she would have joined under full-information. With this structure in place, we then show that any feasible solution that does not have a threshold structure can be perturbed to obtain another feasible solution corresponding to a threshold mechanism with equal or higher throughput. Finally, in Lemma A.1, we show that the set of feasible solutions corresponding to threshold mechanisms forms a compact set under the weak topology. Since the objective of the linear program (2.5) is continuous under the weak topology, we conclude that an optimal signaling mechanism with a threshold structure must exist.

Recall that $M_p \in \mathbb{N}$ is defined such that $h(M_p - 1, p) \geq 0$ and $h(M_p, p) < 0$. Consider any feasible solution $\{\pi_n : n \geq 0\}$ to the linear program (2.5). We first show that we can construct another feasible solution with weakly higher throughput by ensuring the (2.5c) constraints are tight for all $n \leq M_p$. Towards that end, we define $\{\hat{\pi}_n : n \geq 0\}$ by setting

$$\hat{\pi}_n = \begin{cases} \frac{1}{Z}\pi_0 \lambda^n & \text{for } n \leq M_p; \\ \frac{1}{Z}\pi_n & \text{for } n > M_p, \end{cases}$$

where $Z \triangleq \pi_0 \sum_{i=0}^{M_p} \lambda^i + \sum_{i=M_p+1}^{\infty} \pi_i > 0$ is the normalizing constant to ensure $\sum_{n=0}^{\infty} \hat{\pi}_n = 1$.

We first show that $\hat{\pi}$ is feasible for (2.5). From the feasibility of $\{\pi_n : n \geq 0\}$, it is straightforward to show that the constraints (2.5c) continue to hold for $\{\hat{\pi}_n : n \geq 0\}$. Furthermore, we obtain that $\pi_n \leq \pi_0 \lambda^n$ for all $n \geq 0$, and hence $\hat{\pi}_n \geq \pi_n / Z$ for all $n < M_p$. Since $h(n-1, p) \geq 0$ for all $n \leq M_p$, this implies that $\{\hat{\pi}_n : n \geq 0\}$ continues to satisfy the obedience constraint (2.5a). To show that $\hat{\pi}$ is a feasible solution, it remains to verify that (2.5b) holds. For this step, note that we have

$$\begin{aligned} \sum_{n=0}^{\infty} h(n,p)(\lambda \hat{\pi}_n - \hat{\pi}_{n+1}) &= \sum_{n=0}^{M_p - 1} h(n,p)(\lambda \hat{\pi}_n - \hat{\pi}_{n+1}) + \sum_{n=M_p}^{\infty} h(n,p)(\lambda \hat{\pi}_n - \hat{\pi}_{n+1}) \\ &= 0 + \sum_{n=M_p}^{\infty} h(n,p)(\lambda \hat{\pi}_n - \hat{\pi}_{n+1}) \\ &\leq 0. \end{aligned}$$

Here, the second equality follows from the definition of $\hat{\pi}$ and the inequality follows from the fact that $h(n,p) < 0$ for each $n \geq M_p$ and that $\hat{\pi}$ satisfies (2.5c). This proves the feasibility of $\hat{\pi}$.

The difference between the objective values for the two solutions is given by

$$\sum_{n=1}^{\infty} \hat{\pi}_n - \sum_{n=1}^{\infty} \pi_n = \pi_0 - \hat{\pi}_0 = \pi_0 \left(1 - \frac{1}{Z}\right).$$

Now, since $\pi_n \leq \pi_0 \lambda^n$ for all n, we obtain that $Z = \pi_0 \sum_{i=0}^{M_p} \lambda^i + \sum_{i=M_p+1}^{\infty} \pi_i \geq \sum_{n=0}^{\infty} \pi_n = 1$. This implies that the objective value of $\hat{\pi}$ is at least that of π. Furthermore, unless $\pi_n = \lambda^n \pi_0$

for all $n \leq M_p$, we obtain $Z > 1$, implying that the objective value of $\hat{\pi}$ is strictly greater than that of π. From this, we conclude that in any optimal solution π, we must have $\pi_i = \lambda^i \pi_0$ for $i \leq M_p$. Henceforth, we restrict ourselves to feasible solutions π satisfying this property.

Consider now a feasible solution $\pi = \{\pi_n : n \geq 0\}$ such that there exists an $N > M_p$ with $\pi_n = \lambda^n \pi_0$ for all $n < N$, $0 < \pi_N < \lambda \pi_{N-1}$ and $\pi_{N+1} > 0$. We consider a perturbation of this feasible solution and show that it remains feasible and attains the same objective value as the original feasible solution. Toward this goal, define $\tilde{\pi} = \{\tilde{\pi}_n : n \geq 0\}$ as follows

$$\tilde{\pi}_n = \begin{cases} \pi_n & \text{for } n < N; \\ \pi_N + \beta \sum_{i>N} \pi_i & \text{for } n = N; \\ (1-\beta)\pi_n & \text{for } n > N, \end{cases}$$

for some $\beta \in (0, 1]$ to be chosen later. By construction, the LP objective for $\tilde{\pi}$ is same as that for π, and hence it suffices to show that $\tilde{\pi}$ is feasible. Note that $\tilde{\pi}_n \geq 0$ and $\sum_{n=0}^{\infty} \tilde{\pi}_n = 1$, so $\tilde{\pi}$ is a valid distribution. Thus, for $\tilde{\pi}$ to be feasible, we need (2.5a), (2.5b) and (2.5c) to hold.

First, observe that

$$\begin{aligned} \sum_{n=1}^{\infty} \tilde{\pi}_n h(n-1, p) - \sum_{n=1}^{\infty} \pi_n h(n-1, p) &= \beta \sum_{n>N} \pi_n h(N-1, p) - \beta \sum_{n>N} \pi_n h(n-1, p) \\ &= h(N-1, p)\beta \left(\sum_{n>N} \pi_n \left(1 - \frac{h(n-1, p)}{h(N-1, p)} \right) \right) \\ &\geq 0. \end{aligned}$$

The inequality follows from the fact that since $N > M_p$, we have $h(N-1, p) \leq h(M_p, p) < 0$, and the fact that since $h(n, p)$ is non-increasing in n, we have $h(n-1, p)/h(N-1, p) \geq 1$ for $n > N$. Since π is feasible for the LP, this implies that $\tilde{\pi}$ satisfies the constraint (2.5a) for all $\beta \in (0, 1]$.

Next, note that since $\tilde{\pi}_{n+1} = \lambda \tilde{\pi}_n$ for all $n < M_p < N$, we obtain that $\sum_{n=0}^{\infty} h(n, p)(\lambda \tilde{\pi}_n - \tilde{\pi}_{n+1}) = \sum_{n=M_p}^{\infty} h(n, p)(\lambda \tilde{\pi}_n - \tilde{\pi}_{n+1})$. Since $h(n, p) < 0$ for all $n \geq M_p$, the latter expression is non-positive, and (2.5b) holds, if $\lambda\tilde{\pi}_n - \tilde{\pi}_{n+1} \geq 0$ for all n, i.e., if $\tilde{\pi}$ satisfies (2.5c). Finally, it is straightforward to verify that $\tilde{\pi}$ satisfies (2.5c) for all $n \geq 0$ if it is satisfied for $n = N-1$, i.e., if $\lambda\tilde{\pi}_{N-1} - \tilde{\pi}_N \geq 0$. For this condition to hold, we need $\lambda\pi_{N-1} \geq \pi_N + \beta\sum_{i>N}\pi_i$, which holds for any $\beta \in (0,1]$ satisfying $0 < \beta \leq (\lambda\pi_{N-1} - \pi_N)/\sum_{i>N}\pi_i$. So, for any such value of β, we obtain that $\tilde{\pi}$ is feasible for the linear program.

Note that if $(\lambda\pi_{N-1} - \pi_N)/\sum_{i>N}\pi_i \geq 1$, then choosing $\beta = 1$ yields $\tilde{\pi}_n = 0$ for all $n > N$. On the other hand, if $(\lambda\pi_{N-1} - \pi_N)/\sum_{i>N}\pi_i < 1$, then choosing $\beta = (\lambda\pi_{N-1} - \pi_N)/\sum_{i>N}\pi_i$ yields $\lambda\tilde{\pi}_{N-1} - \tilde{\pi}_N = 0$. Thus, we obtain that any $\{\pi_n : n \geq 0\}$, where $\pi_n = \lambda^n\pi_0$ for all $n < N$, $0 < \pi_N < \lambda\pi_{N-1}$ and $\pi_{N+1} > 0$ for some $N > M_p$, can be perturbed appropriately to obtain a feasible solution $\tilde{\pi}$ with equal objective and satisfying either (1) $\tilde{\pi}_n = \lambda^n\tilde{\pi}_0$ for all $n < N$, $0 < \tilde{\pi}_N \leq \lambda\tilde{\pi}_{N-1}$, and $\tilde{\pi}_n = 0$ for all $n > N$ or (2) $\tilde{\pi}_n = \lambda^n\tilde{\pi}_0$ for all $n \leq N$. In the latter case, if $0 < \tilde{\pi}_{N+1} < \lambda\tilde{\pi}_N$, one can perturb $\tilde{\pi}$ analogously. By induction, this implies that if the LP optimum is attained, then it is attained by a feasible solution $\{\pi_n : n \geq 0\}$ for which there exists an $N \geq M_p$ with $\pi_n = \lambda^n\pi_0$ for all $n < N$, $0 < \pi_N \leq \lambda\pi_{N-1}$ and $\pi_n = 0$ for all $n > N$. (Here, N could be infinite.) Hence, we restrict our attention to feasible solutions of this form.

In Lemma A.1, we show that the set of all such feasible distributions is compact (under the weak topology). Since the objective is a continuous function of the steady state distribution, we obtain that an optimal solution of this form exists.

Summarizing, there exists an optimal solution $\{\pi_n : n \geq 0\}$ to the LP (2.5) for which there exists an $N \geq M_p$ (possibly infinity) such that $\pi_n = \lambda^n\pi_0$ for all $n < N$, $0 < \pi_N \leq \lambda\pi_{N-1}$ and $\pi_n = 0$ for all $n > N$. Finally, using Lemma 2.3, this implies that there exists an optimal signaling mechanism σ for which there exists an $N \geq M_p$ and a $q = \pi_N/(\lambda\pi_{N-1}) \in [0,1]$ such that $\sigma(n,1) = 1$ for all $n < N$, $\sigma(N,1) = q$ and $\sigma(n,1) = 0$ for all $n > N$. If $N = \infty$, we obtain that the mechanism $\sigma = \sigma^\infty$ is optimal. Otherwise, we obtain that σ^{N+q} is an optimal signaling mechanism. □

The preceding theorem has an important practical implication: the optimal signaling mechanism is easy to describe and implement. More precisely, our analysis assumes that the service provider can publicly announce and commit to a signaling mechanism. Given that signaling mechanisms are arbitrary mappings over the set of all non-negative integers, this is a strong assumption for general signaling mechanisms. However, the structure of the optimal signaling mechanism renders this assumption innocuous. In particular, the service provider can easily implement the signaling mechanism σ^x by announcing *a priori* the value $N = \lfloor x \rfloor$ below which customers will be deterministically recommended to join the queue, and the probability $q = x - \lfloor x \rfloor$ with which they will be recommended to join when the queue length is exactly N.

We note that our proof implies that in any optimal signaling mechanism (not necessarily threshold), no customer would be told to leave if they would have joined under full information. This follows from the fact in any optimal solution π to (2.5), we have $\pi_{n+1} = \lambda \pi_n$ for all $n < M_p$. Notice that whenever a customer is told to leave, they know the length of the queue is more than M_p and joining will get them negative utility. This is similar to the results of Kamenica and Gentzkow (2011) where they showed that in an optimal persuasion mechanism, whenever an agent is told to take the principal's least-preferred action, the agent knows with certainty that it is in her best interest.

2.5 Optimal pricing and signaling

Having determined the structure of the optimal signaling mechanism for any fixed price p, we next investigate the service provider's decision problem of how to set p optimally in order to maximize her revenue.

In order to understand this problem, consider first as a detour the case of optimal state-dependent pricing in a fully-observable queue. More precisely, consider the setting where the arriving customers can observe the queue-length, but the service provider is allowed to charge them a price dependent on the queue-length. This setting of dynamic pricing serves as a natural benchmark against which we compare the revenue obtained under the optimal fixed-price and signaling mechanism. Surprisingly, we find that setting the fixed price optimally along with using an optimal signaling mechanism suffices to achieve the optimal revenue in the observable setting. We have the following main theorem.

Theorem 2.2. *With the optimal value of the fixed price p and the corresponding optimal signaling mechanism, the service provider obtains the same revenue as under optimal state-dependent prices in a fully observable queue.*

Before we state the proof, we note an important practical implication of this result. In many settings, state-dependent pricing is infeasible, either due to operational reasons, such as price stickiness arising out of menu costs associated with changing prices (Sheshinski and Weiss 1977), or due to exogenous reasons such maintaining customer expectations about the price of service (Kalwani et al. 1990). In such settings, however, it may be feasible to make recommendations to customers based on the state of the system. Our result states that in such settings, as long as the fixed-price is chosen optimally, the service provider can effectively use signaling to guarantee the same optimal revenue as with optimal state-dependent pricing.

Proof of Theorem 2.2. Let the optimal state-dependent pricing mechanism set a price $p(n)$ for service to an arriving customer when the number of customers already in queue is n. Under our assumption of non-increasing utilities, the prices $\{p(n) : n \geq 0\}$ can be shown to have the following form (Chen and Frank 2001): up to a threshold of the queue length, the service provider sets prices that extract out all the surplus of the incoming customer, making those customers indifferent between joining and leaving; beyond this threshold, the service provider sets a large price, essentially denying entry to any incoming customers. Formally, the optimal prices satisfy $p(n) = u(n)$ for all $n < \kappa$ and $p(n) = \infty$ for all $n \geq \kappa$, for an appropriately chosen $\kappa > 0$. Let π_∞^κ denote the steady state distribution of the queue under this pricing policy, and let X_∞^κ denote an (independent) random variable distributed as π_∞^κ. Note that under optimal state-dependent prices, the service provider's expected revenue is given by $\lambda\,\mathbf{E}\left[\mathbf{I}\{X_\infty^\kappa < \kappa\}p(X_\infty^\kappa)\right] = \lambda\,\mathbf{E}\left[\mathbf{I}\{X_\infty^\kappa < \kappa\}u(X_\infty^\kappa)\right]$.

Now, for the setting of an unobservable queue, consider the fixed price $\hat{p} \triangleq \mathbf{E}\left[u(X_\infty^\kappa)|X_\infty^\kappa < \kappa\right]$, and threshold signaling mechanism σ^κ, i.e., the service provider sends signal 1 (or "join") if the queue length is strictly less than κ, and 0 (or "leave") otherwise. We claim that with this choice of fixed price $\hat{p}$ and the signaling mechanism σ^κ, the service provider achieves the same expected revenue in the obedient equilibrium as under the optimal state-dependent mechanism.

We start by showing that under the fixed price $\hat{p}$ and the threshold signaling mechanism σ^κ, the obedient strategy forms a customer equilibrium. To see this, observe that if all customers follow the recommendation, the steady state distribution of the queue-length is indeed given by π_∞^κ. By an abuse of notation, we let X_∞^κ denote the queue-length upon a particular customer's arrival. Thus, the expected payoff to the customer for joining the queue upon receiving the signal $s = 1$ is given by

$$\mathbf{E}[h(X_\infty^\kappa, \hat{p})|s = 1] = \mathbf{E}[u(X_\infty^\kappa)|s = 1] - \hat{p} = \mathbf{E}[u(X_\infty^\kappa)|X_\infty^\kappa < \kappa] - \hat{p} = 0,$$

where the second equality follows from the fact that the signaling mechanism σ^k sends signal $s = 1$ if and only if $X_\infty^\kappa < \kappa$. This implies that the resulting steady state distribution satisfies the first obedience constraint (2.5a). Similarly, the expected payoff to the customer upon receiving the signal $s = 0$ is given by

$$\mathbf{E}[h(X_\infty^\kappa, \hat{p})|s = 0] = \mathbf{E}[u(X_\infty^\kappa)|s = 0] - \hat{p} = u(\kappa) - \hat{p} = u(\kappa) - \mathbf{E}[u(X_\infty^\kappa)|X_\infty^\kappa < \kappa] \leq 0,$$

where the second equality follows from the fact under the steady state the signaling mechanism sends signal $s = 0$ if and only if $X_\infty^\kappa = \kappa$, the third equality from the definition of $\hat{p}$, and the inequality holds because u is non-increasing. This implies that the resulting steady state distribution satisfies the second obedience constraint (2.5b).

Next, observe that the service provider's expected revenue is given by

$$\lambda\,\mathbf{E}[\sigma(X_\infty^\kappa,1)\hat{p}] = \lambda\,\mathbf{E}\left[\mathbf{I}\{X_\infty^\kappa < \kappa\}\hat{p}\right] = \lambda\,\mathbf{E}\left[\mathbf{I}\{X_\infty^\kappa < \kappa\}\,\mathbf{E}\left[u(X_\infty^\kappa)|X_\infty^\kappa < \kappa\right]\right] = \lambda\,\mathbf{E}\left[\mathbf{I}\{X_\infty^\kappa < \kappa\}u(X_\infty^\kappa)\right],$$

where the last equality follows from the tower property of conditional expectation. Thus, we observe that the service provider's expected revenue is same as that of the optimal state-dependent pricing mechanism.

Finally, with homogeneous customers, the optimal state-dependent pricing mechanism is welfare-maximizing (Chen and Frank 2001) with zero customer surplus. Since the optimal fixed price $\hat{p}$ and the signaling mechanism σ^κ achieve this revenue, these values must be optimal. Thus, we obtain that with the optimal choice of the fixed price and the corresponding optimal signaling mechanism, the service provider obtains the same revenue as the optimal state-dependent prices. □

The preceding theorem leads to a natural question: can the service provider increase her revenue in an unobservable queue through a combination of signaling and pricing? Under such a mechanism, customers not only receive information about the queue state from the signal but also from the price. Despite this generality, our result already implies that such mechanisms cannot improve the revenue. This follows from the fact that, in a fully observable queue, the optimal state-dependent pricing mechanism is welfare-maximizing and has zero customer surplus (Chen and Frank 2001). Since customer surplus in any mechanism must be non-negative, there cannot be any combination of signaling and pricing that achieves strictly higher revenue than the optimal state-dependent pricing mechanism (or the optimal signaling mechanism with an optimal fixed-price). However, there exist many mechanisms that achieve this optimal revenue through a combination of signaling and pricing. Specifically, given any partition of the set of queue-lengths for which a customer joins under the state-dependent pricing mechanism, one can construct a combined signaling-and-pricing mechanism that achieves the optimal revenue: such a mechanism would reveal to an arriving customer which set of the partition the queue-length lies in, and charge them the expected utility conditioned on the queue-length being in that set. Furthermore, as long the prices for different sets of the partition are different, the prices can themselves act as signals. This discussion suggests that, in an unobservable queue, the service provider has a flexibility in choosing the number of prices while optimizing her revenue.

2.6 Extensions

In this section, we discuss a few extensions to our results and our model. First, for the special case where the customers utility is linear in time spent in queue, we obtain a closed-form expression for the threshold in the optimal signaling mechanism as a function of the fixed-price p. Subsequently, we discuss how our model can be extended to include abandonment and customer heterogeneity.

2.6.1 Linear utility

A commonly studied model for customer utility is one where the customer receives a fixed value $V > 0$ from service, and incurs a disutility that is proportional to the time spent while waiting until service completion.(Naor 1969, Allon et al. 2011, Borgs et al. 2014) Since we assume that the customers' service requirements are homogeneous and have unit mean, this assumption implies that the customer utility $u(\cdot)$ is given by $u(X) = V - c(X + 1)$ for all $X \geq 0$, for some value of $c > 0$ that denotes the disutility per unit time of waiting. For this utility model, Theorem 2.1 implies that the linear program 2.5 can be analytically solved, resulting in a closed-form expression for the threshold in the optimal signaling policy.

To state our results, we assume, without loss of generality, that $V = 1$, and, to avoid trivialities, we let $c \in (0, 1)$. Furthermore, let $W_0(\cdot)$ and $W_{-1}(\cdot)$ denote the two real branches of the Lambert-W function, defined as the set of functions that are the inverse of $f(X) = Xe^X$. (See Borgs et al. (2014) for a detailed description.) We have the following theorem.

Theorem 2.3. *Suppose $u(n) = 1 - c(n + 1)$ with $c \in (0, 1)$. Then, for each $p \in [0, 1 - c]$, the threshold mechanism σ^x is optimal for $x = N + q$, where*

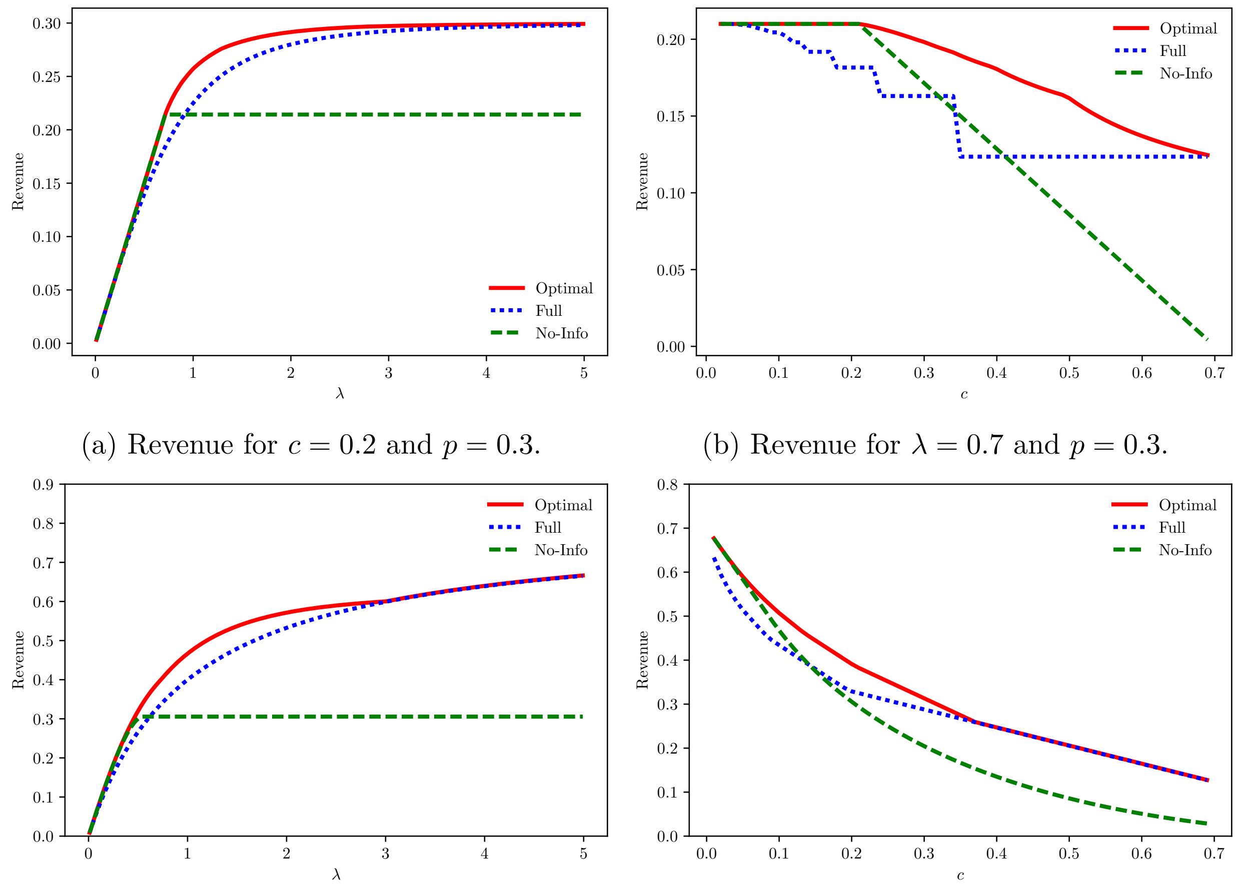
$$N = \begin{cases} \left\lfloor \frac{2(1-p)}{c} - 1 \right\rfloor & \text{if } \lambda = 1; \\ \infty & \text{if } \lambda \leq 1 - \frac{c}{1-p}; \\ \left\lfloor \frac{1}{\log(\lambda)} \left(W_i\left(-\kappa e^{-\kappa}\right) + \kappa \right) \right\rfloor & \text{otherwise,} \end{cases}$$

with $\kappa = \left(\frac{1-p}{c} - \frac{1}{1-\lambda}\right) \log(\lambda)$ and where $i = 0$ when $\lambda > 1$ and $i = -1$ when $1 - \frac{c}{1-p} < \lambda < 1$. For all values of $\lambda < \infty$, we have

$$q = \frac{\sum_{k<N} \lambda^k (1 - p - c(k + 1))}{\lambda^N (c(N + 1) + p - 1)} \in [0, 1].$$

The proof involves first showing that the throughput is increasing in the threshold as long as the obedient strategy is a customer equilibrium for the corresponding threshold mechanism. Then, using the equilibrium conditions for the obedient equilibrium, we obtain bounds on the optimal thresholds. We provide the full details in Appendix A.1.1.

Using this closed form expression, we numerically compare the optimal signaling mechanism against those of fully-revealing and no-information mechanisms. In Figure 2.1a, we plot the revenue of the optimal mechanism, along with those of the fully-revealing and the no-information mechanisms for a range of values of λ, when the customer utility is given by $u(X) = 1 - c(X + 1)$ with $c = 0.2$ under a fixed price $p = 0.3$. As λ increases, the revenue of the fully-revealing mechanism and the optimal mechanism both converge to 0.3, the value when throughput is equal to 1; however, for any fixed λ, the optimal mechanism outperforms the others. Note that, for small arrival rates, the no-information mechanism outperforms



(a) Revenue for $c = 0.2$ and $p = 0.3$.

(b) Revenue for $\lambda = 0.7$ and $p = 0.3$.

(c) Optimal revenue for $c = 0.2$.

(d) Optimal revenue for $\lambda = 0.7$.

Figure 2.1: Comparison of the optimal, the fully-revealing, and the no-information mechanisms.

the fully-revealing one; this is consistent with existing results (Simhon et al. 2016), as in this region, under no-information mechanism, all customers join the queue, whereas under fully-revealing mechanism, some customers will not join upon seeing a long queue.

Next, we consider the effect of changing c while fixing the values of the arrival rate ($\lambda = 0.7$) and the fixed-price ($p = 0.3$) in Figure 2.1b. Notice the revenue for the fully-revealing mechanism is discontinuous: this is because under full-information, customers join the queue only if the queue length is strictly less than $\frac{1-p-c}{c}$, implying that the customer strategy is discontinuous in c. For low values of c, the revenue of the no-information and the optimal mechanism are both equal to $p\lambda = 0.21$, the maximal value, since all customers join the queue. As c increases, the revenue of the no-information mechanism goes to zero, whereas the revenue of fully-revealing and the optimal mechanism goes to 0.4, which arises when the customers only join the queue if it is empty when they arrive.

Finally, in Figures 2.1c and 2.1d, we compare the revenue under the optimal mechanism against the revenue of the no-information and the fully-revealing mechanisms, where under each setting we set the fixed-price optimally. In particular, in Figure 2.1c, we fix $c = 0.2$ and vary λ, whereas in Figure 2.1d, we fix $\lambda = 0.7$ and vary c. We see that the no-information outperforms the fully-revealing information for low values of the arrival rate λ, for a given c. As λ increases, we observe that the revenue of the optimal and the fully-revealing mechanisms converge, while the no-information mechanism's revenue is much lower. Similarly, for a fixed arrival rate, we see that optimal signaling is effective in increasing revenue over the no-information and the fully-revealing mechanisms for moderate values of c, i.e., when customers incur moderate disutility for waiting.

2.6.2 Abandonment

Our model assumes that the customers who choose to join the queue stay in queue until service completion. In many settings, this modeling assumption is unrealistic, and one must explicitly account for customer abandonment. A standard approach (Garnett et al. 2002) to incorporate customer abandonment is by modeling each arriving customer to have an independent and exogenously specified patience time τ, and assuming that the customer will abandon the queue if she is still waiting to be served at time τ after joining the queue.

Consider the setting where customers' patience times are distributed independently and identically as an exponential distribution with rate γ. It is straightforward to show that our results continue to hold in this setting. Formally, as before, let $h(n, p)$ denote a customers' payoff upon joining the queue with n customers already in queue, and the fixed price is p.

(Note that unlike in our original model, this payoff function now incorporates the fact that the customer may leave without obtaining service.) Then, by a similar argument as before, the service provider's decision problem can be written as the following linear program:

$$\begin{aligned}
&\max_{\pi} \quad \sum_{n=1}^{\infty} \pi_n \\
&\text{subject to,} \quad \sum_{n=0}^{\infty} (1+\gamma n)\pi_{n+1} h(n,p) \geq 0, \\
&\sum_{n=0}^{\infty} \left(\lambda \pi_n - (1+\gamma n)\pi_{n+1}\right) h(n,p) \leq 0, \\
&\lambda \pi_n - (1+\gamma n)\pi_{n+1} \geq 0, \quad \text{for all } n \geq 0, \\
&\sum_{n=0}^{\infty} \pi_n = 1, \qquad \pi_n \geq 0, \quad \text{for all } n \geq 0.
\end{aligned}$$

(Here, we assume that once a customer is in service, she would not abandon the queue; without this assumption, one obtains a slightly modified linear program.) From each feasible solution π to this linear program, one can obtain the corresponding signaling mechanism σ as $\sigma(n,1) = \frac{(1+\gamma n)\pi_{n+1}}{\lambda \pi_n}$. For this setting, a similar analysis of the preceding linear program establishes Theorem 2.1 under same monotonicity conditions on the payoff function h.

Note however that the preceding model assumes that the customers only choose to abandon the queue when their patience runs out, and never before. When the queue is observable and the service time distributions are known, this is a fairly mild assumption, since a customer does not learn new information about her waiting time while she waits in the queue. However, in an unobservable queue, this assumption is strong and will in fact not be followed by a fully rational customer. In particular, a rational customer may find it optimal to abandon the queue before her patience runs out. This is because, since the time spent waiting in queue provides further information to a customer regarding the queue length, it is rational for a customer to abandon the queue if she believes that her waiting might be larger than her remaining patience. Modeling the abandonment decision endogenously is challenging even in models without signaling (Ata and Peng 2017, Ata et al. 2017), and incorporating signaling in such models is an interesting direction for future work.

2.6.3 Heterogeneous types

Another assumption of our model is that the customers are homogeneous. Our model can be naturally extended to allow for customer heterogeneity in the form of different utility for joining the queue. Formally, suppose there are K possible customer types $i \in \{1, \cdots, K\}$, where the arrival rate of customers of type i is given by λ_i. Suppose also that the service provider observes the type of a customer upon her arrival, and charges a customer of type i a fixed-price p_i for obtaining service. Let $\sum_{i=1}^{K} \lambda_i = \Lambda$. Denote the expected payoff for a customer of type i upon joining the queue with X customers already in queue as $h_i(X, p_i) \triangleq u_i(X) - p_i$, where each $u_i(\cdot)$ is non-increasing with $u_i(0) > 0$ and $\lim_{X\to\infty} u_i(X) < 0$. In this setting, a signaling mechanism is specified by $\{\sigma(n, i, j) : n \geq 0, i = 1, \cdots, K; j = 0, 1\}$, where $\sigma(n, i, 1)$ denotes the probability with which the service provider tells a customer of type i to join the queue when there are n customers already in queue and $\sigma(n, i, 0)$ denotes the probability of telling them not to join. By letting $\phi_n^{i,j} = \frac{\lambda_i}{\Lambda}\pi_n \sigma(n, i, j)$, where $\pi = \{\pi_n : n \geq 0\}$ is the steady-state distribution of the queue, the service provider's decision problem can be reduced to the following linear program:

$$
\begin{aligned}
\max_{\phi} \quad & \sum_{n=0}^{\infty}\sum_{i=1}^{K} p_i \cdot \phi_n^{i,1} & \\
\text{subject to,} \quad & \sum_{n=0}^{\infty} \phi_n^{i,1} h_i(n, p_i) \geq 0, & \text{for } i = 1, 2, \ldots, K, \\
& \sum_{n=0}^{\infty} \phi_n^{i,0} h_i(n, p_i) \leq 0, & \text{for } i = 1, 2, \ldots, K, \\
& \frac{1}{\lambda_1}\left(\phi_n^{1,0} + \phi_n^{1,1}\right) = \frac{1}{\lambda_i}\left(\phi_n^{i,0} + \phi_n^{i,1}\right), & \text{for all } n \geq 0,\ i = 2, 3 \ldots, K, \\
& \frac{1}{\lambda_1}\left(\phi_n^{1,0} + \phi_n^{1,1}\right) = \sum_{i=1}^{K} \phi_{n-1}^{i,1} & \text{for all } n \geq 1, \\
& \sum_{n=0}^{\infty}\sum_{i=1}^{K} \phi_n^{i,1} + \phi_n^{i,0} = 1 & \\
& \phi_n^{i,j} \geq 0 & \text{for all } n \geq 0,\ i = 1, 2, \ldots, K, \text{ and } j = 0, 1.
\end{aligned}
\tag{2.6}
$$

Here, the first two inequalities correspond to the obedience constraints for each type of customer. In particular, the first inequality requires that when a type i customer is told to join the queue, she finds it optimal to join, whereas the second inequality requires that the customer finds it optimal not to join the queue if she is told not to join. The remaining constraints in the linear program arise from the constraints on the steady-state distribution π. From a feasible solution ϕ, one can obtain the signaling mechanism as $\sigma(n,i,1) = \frac{\phi_n^{i,1}}{\phi_n^{i,1}+\phi_n^{i,0}}$ and $\sigma(n,i,0) = \frac{\phi_n^{i,0}}{\phi_n^{i,1}+\phi_n^{i,0}}$. Note that if $K = 1$, we are back to the case with homogeneous customers, and the preceding linear program reduces to the linear program (2.5), where $\phi_n^{1,1} = \pi_{n+1}/\lambda$, and $\phi_n^{1,0} = \pi_n - \pi_{n+1}/\lambda$ for all n.

For the heterogeneous customer type setting, our main result, Theorem 2.1, extends as follows:

Theorem 2.4. *Suppose all the customer types are charged the same fixed-price, i.e., $p_i = p$ for all $i = 1, \ldots, K$. Then, there exists an optimal signaling mechanism that signals each customer type using a threshold mechanism: $\sigma(n,i,1) = 1$ for $n < N_i$ and $\sigma(n,i,1) = 0$ for $n > N_i$ for some N_i.*

This result is obtained using a similar argument as to our main result: first, we show that under the optimal mechanism, each customer type is told to join the queue at all queue-lengths for which they would have joined under full-information; next, we show that any feasible mechanism satisfying this property but not of a threshold type can be perturbed appropriately without reducing the revenue. We omit the details for brevity.

On the other hand, if not all customer types are charged the same price, threshold mechanisms need not be revenue-optimal across all signaling mechanisms. We illustrate this using the following example: consider a setting with two types of customers ($K = 2$), where each customer type has the following utility function:

$$u_1(n) = \begin{cases} 51 & n = 0 \\ 40 & n = 1 \\ -10,000 & n \geq 2 \end{cases}, \qquad u_2(n) = \begin{cases} 2 & n \leq 1 \\ -8.5 & n \geq 2 \end{cases}.$$

The type 1 customers are charged a price $p_1 = 50$ for service, whereas the type 2 customers are charged $p_2 = 1$. The arrival rates of the two types are $\lambda_1 = \lambda_2 = 1$. Solving the linear program (2.6), we obtain the optimal signaling mechanism to be

$$\sigma(n, 1, 1) = \begin{cases} 1 & n = 0 \\ \frac{1}{10} & n = 1 \\ 0 & n \geq 2 \end{cases}, \qquad \sigma(n, 2, 1) = \begin{cases} 0 & n = 0 \\ \frac{1}{10} & n = 1 \\ 0 & n \geq 2 \end{cases}.$$

Observe that this is not a threshold mechanism for customers of type 2. One can verify that no threshold mechanism achieves the same revenue as the preceding mechanism.

Nevertheless, the following theorem shows that if the (fixed) price for each type is set optimally, the revenue-optimal signaling mechanism is a threshold mechanism. Furthermore, an analogous result as in Theorem 2.2 holds: the optimal signaling mechanism (together with optimally set fixed type-dependent prices) achieves the same revenue as the optimal state-and-type-dependent pricing mechanism. We provide the proof in Appendix A.1.2.

Theorem 2.5. *For the optimal choice of fixed prices $p_i, i = 1, \cdots, K$, the optimal signaling mechanism has a threshold structure. Also, the revenue achieved by the service provider under this mechanism is same as that in the optimal state-and-type dependent pricing mechanism.*

Finally, a further extension of our model to heterogeneous customers involves the setting of private types, where the service provider cannot observe the types of the arriving customers. These settings in general require a combinatorial number of signals, where each signal corresponds to a subset of customer types who join the queue after receiving it. In the special case where all customer types are charged the same price p, and the types are ordered, meaning $h_i(n,p) \geq h_{i+1}(n,p)$ for all $i = 1, \ldots, K$ and $n \geq 0$, it suffices to consider mechanisms involving $K+1$ signals, where the signal i corresponds to all customers with types less than or equal to i joining the queue. Although we can again formulate the service provider's decision problem as a linear program, we note once again that threshold mechanisms need not be optimal; there may exist signaling mechanisms that obtain higher revenue than threshold mechanisms.

2.7 Conclusion

We analyze optimal information design in the context of an unobservable $M/M/1$ queue with strategic customers. We first establish that in the optimal signaling mechanism, the service provider does not fully reveal the queue state, nor completely conceals it: instead, the optimal signaling mechanism uses binary signals in a threshold structure. Further, we show that with the optimal choice of the fixed price, the service provider can use signaling to achieve the expected revenue achieved by the optimal state-dependent pricing mechanism.

Throughout this (and the next chapter), we restrict to customers who seek to maximize some expected utility. Many customers are ill-modeled by expected utility: for example, people exhibit caution over the tails of the wait-time distribution (Maister et al. 1984). In Chapter 4, we consider customers who have more general utilities. We conclude that chapter by applying its methods to this chapter's queueing model, and, as an example, consider a customer whose disutility from waiting is their expected waiting time plus its variance.

We now proceed with another application of information design. In Chapter 3, we consider persuasion in online retail, and use a similar linear programming approach to the proof of Theorem 2.1 to find the revenue-optimal approach to signaling.

CHAPTER 3

PERSUASION IN ONLINE RETAIL: EFFICACY OF PUBLIC SIGNALS

How come anything you buy will
go on sale next week?

Erma Bombeck

3.1 Introduction

With the rapid growth of online retail, a number of auxiliary services have arisen that provide customers with information about price history, product availability, historical demand, etc., to help them in the search for a better price. For example, services like Kayak inform a user whether the price of an airline ticket is likely to fall in the future. Whenever a customer using such services waits for the price to decrease, without direct access to demand and inventory information, she risks the product selling out. To alleviate this risk and to induce customers to purchase earlier at a greater price, online retailers often provide customers with demand and inventory information. Some retailers, such as Amazon, give a count of the items remaining when the inventory is low, while others display a "low stock" indicator. This raises a natural question: how should a retailer credibly communicate inventory and demand information to customers to maximize its expected revenue?

To study this question, we consider the setting where a retailer seeks to sell their inventory of an item that will soon drop in price. Common examples of this are fashion (where summer wear is sold in mass clearance sales after autumn arrives) and theater tickets (where the price of unsold tickets drops right before a show). In particular, we consider a two period model: prices are high the first time period and low the second, and customers who arrive at time 1 decide whether to buy the item right away or wait to buy in the next time period. The firm sends each customer at time 1 a signal that depends on both the total inventory and the total demand at time 1. To concentrate on the effect of signaling, we let the prices be determined exogenously. When a manufacturer sets price restrictions, this can be seen as the only lever available to an online retail platform like Amazon or Ebay. Each customer is strategic and Bayesian: before they act, they update their beliefs about the inventory and the demand based on the firm's signal.

In general, the firm may send a *private* signal to each customer, possibly giving them different information. For example, the firm may send a tailored email message to each interested customer. However, such private signaling may not always be feasible, or may be hard to implement. Furthermore, due to lack of transparency, customers might feel slighted for not receiving valuable information given to other customers. Relatedly, private signals are also susceptible to "leakage," where customers share their signals with each other, leading to unanticipated effects. Given these issues, a firm may consider sharing information *publicly*, where all customers receive the same signal. For example, the firm may put a "low stock" indicator on its website visible to all interested customers. While public signals are transparent, equitable, and leak-proof, it is unclear whether they are equally effective in raising the firm's revenue. Furthermore, from a computational standpoint, it is NP-hard in general to obtain even a constant factor approximation of the optimal public mechanism (Dughmi and Xu 2017).

Our primary result is that with homogeneous customers, the optimal signaling mechanism is indeed *public*. The optimal mechanism sends all customers a common binary signal (e.g., "buy now" and "wait", or "in-stock" and "low-stock") recommending all of them to take the same action. We establish this result by first formulating the firm's problem as an LP, and showing that any feasible solution can be altered to a public one without decreasing the revenue. The latter result relies on two intermediate lemmas characterizing the nature of competition among the customers. Furthermore, using this result, we show that the problem of finding the optimal signaling mechanism can be posed as a fractional knapsack problem, yielding an efficient linear time algorithm.

Our numerical investigations show that the optimal signaling mechanism achieves a substantial increase in revenue, often getting close to the setting where all time 1 customers are compelled to buy immediately (i.e., where there is no clearance period). Although our analytical result does not extend in the presence of customer heterogeneity, we demonstrate numerically that simple public signals continue to obtain a significant fraction of the revenue achieved through private signaling, with their performance improving when the customers become increasingly differentiated. (We discuss results for customer heterogeneity in Section 3.5.)

Our analysis assumes that the firm *commits* to sharing information in a way it chooses in advance. While the question of how a firm may achieve such commitment power is beyond our scope, we provide two justification for this assumption. First, without credible information sharing, the firm cannot increase its revenue beyond the setting of no information sharing (Allon and Bassamboo 2011): if a firm says all items have low inventory, the customers will simply ignore the warning altogether. Because even committing to share full information often achieves higher revenue than no information sharing, a firm gains significantly from *committing* even to a suboptimal mechanism. Second, from a practical perspective, retailers are typically loath to mislead customers due to potential reputation loss. In contrast, by committing, they stand to gain through trust and reputation effects.

3.1.1 Literature review

Inventory communication in online retail has received substantial attention, both from a theoretical (Allon and Bassamboo 2011, Aydinliyim et al. 2017, Cui and Shin 2018) and empirical perspective (Peinkofer et al. 2016, Cui et al. 2018). Our work builds on the two-period model of Allon and Bassamboo (2011), who analyze it under "cheap-talk" i.e., the firm cannot commit to a signaling mechanism. The authors show the retailer cannot increase revenue over the no-information mechanism. As mentioned earlier, we study signaling with commitment and show that the firm can obtain a substantial increase in revenue from public mechanisms. Aydinliyim et al. (2017) consider signaling and pricing in a two-period model with commitment, but restrict the signaling mechanisms to masking (simply stating "in stock") or sharing the exact inventory. They find sufficient conditions for masking or sharing being the optimal choice.

Contemporary and independent to our work, Drakopoulos et al. (2018) study pricing and signaling under commitment in a two-period model with a (fixed) continuum of heterogeneous customers, where the inventory takes one of two values. Specifically, under high stock, all customer demand is met, whereas under low stock, a known fraction of the demand can be satisfied. The authors study the interplay between pricing and signaling in this setting and find that private signaling outperforms sharing information publicly. In contrast, we consider a model with homogeneous customers, fixed prices, and an arbitrary (joint) distribution of inventory and demand, and find that public signaling mechanisms are optimal. Moreover, in our setting the customers' beliefs are *size-biased* due to the demand being unknown to the customers, a phenomenon observed in other settings with a random number of agents (McAfee and McMillan 1987). Aviv et al. (2018) consider how responsive pricing can impact a two-period retail model with heterogeneous customers. When customers are myopic, they find that responsive pricing leads to higher revenue. However, when customers are strategic, they find that frequently, fixed announced prices perform better. Jiang et al. (2016) consider information sharing between a retailer and a market-demand observing manufacturer. They show the retailer prefers no information be shared, since this forces the manufacturer to charge lower wholesale prices when demand is low. This is unlike our results, where because information is the only lever available to the firm, the uninformed customer prefers any information about demand and supply over none. Cui et al. (2018) empirically show the effect of the demand and supply signaling we consider in this chapter. They show this by generating increased revenue when they manipulate Amazon.com's supply information to indicate lower stock and higher quality.

Our work adopts the approach of Bayesian persuasion (Kamenica and Gentzkow 2011, Bergemann and Morris 2018). Recent work has applied this approach to study crowdsourcing (Papanastasiou et al. 2017, Kremer et al. 2014), queueing (Lingenbrink and Iyer 2018a), content sharing in networks (Candogan and Drakopoulos 2017) and disaster management (Alizamir et al. 2018). Our work extends this literature by assuming a random agent population, which causes agents' beliefs to be *size-biased* towards larger values of demand. Rayo and Segal (2010) consider a setting where a sender chooses how to signal a prospect, which has two parameters: the profit to the sender and the relevance to a Bayesian heterogeneous receiver. Unlike our model, receivers have no externalities for their decisions, and choose to interact with the prospect if they believe the relevance to be above a threshold. Like our results, they find that the sender's preferred signaling will pool profitable prospects with less profitable but more relevant prospects.

Taneva (2019) considers a game with two homogeneous agents and one principal where the agents play a two-state two-action game with payoffs dependent on the unknown state. The principal can send a signal to the players to convey the state. She finds settings where public signaling performs worse than private signaling, despite the homogeneity of the agents. The homogeneity of our model, then, does not immediately suggest public signaling's optimality. Dughmi and Xu (2017) study signaling in a game with no externalities: a customer's utility is not dependent on the actions of the other customers. In the case of private signaling, they find a $(1 - 1/e)$-approximation algorithm for the optimal private signaling mechanism. For public signaling, they show it is, in general, NP-hard to approximate the optimal public signaling mechanism by any constant factor. In contrast to this work, in our model, there are externalities among the customers: if more customers buy earlier, then a customer's utility for waiting decreases. They also allow for customer heterogeneity in their utility, while we restrict customer utility to be homogeneous until we relax this in Section 3.5. For further information on approximation of optimal public and private signaling, we refer the reader to the survey Dughmi (2017).

3.2 Model

We consider a two-period retailing model similar to the one in Allon and Bassamboo (2011), where a single firm seeks to sell its limited supply of a product over periods $t = 1, 2$. The total inventory of the product (at time 1) is given by Q_1, and there is no new inventory arriving at time 2. The price of the product at time t is exogenously fixed at $p_t > 0$. We are interested in a setting where the second period acts as a clearance or a sale period, in which the firm seeks to liquidate its total inventory. Consequently, we assume the price in the second period is lower, i.e., $p_1 > p_2$[1].

The firm sees the arrival of N_t *customers* at time $t \in \{1, 2\}$. The customers who arrive at time $t = 1$ share a value $v > p_1$ for the product[2]; the potential customers who arrive at time $t = 2$ may have different valuations of the product, but all value the product over the second period price, p_2. We call the customers who attempt to purchase the item *interested customers*, and let D_t denote the number of interested customers at time $t \in \{1, 2\}$. Because of all customers value the item over the second period price, all customers still in the system at time $t = 2$ will be customers at time $t = 2$. On the other hand, customers at time 1 may choose to delay their purchase to time 2 at a waiting cost of $c \geq 0$.

At any time t, if there are more interested customers than there are items available, the firm uniformly selects a random subset to sell all the inventory. Letting Q_2 denote the

[1]Note that if $p_1 \leq p_2$, a customer's optimal action is to buy at time 1 regardless of the quantity or the demand. Thus signaling cannot raise the revenue in this setting.

[2]We consider customer heterogeneity in Section 3.5.

remaining inventory at time 2, we have $D_2 = N_1 + N_2 - D_1$ and $Q_2 = (Q_1 - D_1)^+$. Then, the probability an interested customer at time t receives the product is given by $\min\left\{\frac{Q_t}{D_t}, 1\right\}$. Since the decision of the customers at time 2 is trivial, we henceforth let the term "customers" refer to those at time 1 whenever the context is clear.

The firm's revenue is simply $R(Q_1, N_1, N_2, D_1) \triangleq \sum_{i=1,2} p_i \min\{Q_i, D_i\}$, which can be simplified to

$$R(Q_1, N_1, N_2, D_1) \triangleq p_2 \min\{Q_1, N_1 + N_2\} + (p_1 - p_2) \min\{Q_1, D_1\}.$$

To describe customers' utility more precisely, we fix a customer at time 1, and let $\widehat{D} \leq N_1 - 1$ denote the number of *other* interested customers at time 1. Then, if the customer seeks to buy at time 1, her utility is given by $u_1(Q_1, \widehat{D}) = \min\left\{\frac{Q_1}{\widehat{D}+1}, 1\right\}(v - p_1) > 0$. On the other hand, if the customer chooses to wait until time 2, her utility is given by $u_2(Q_1, N_1, N_2, \widehat{D}) = \min\left\{\frac{(Q_1 - \widehat{D})^+}{N_1 - \widehat{D} + N_2}, 1\right\}(v - p_2) - c$. Given this basic model, we now describe the belief structure and the decision problems of the firm and the customers.

Information structure: There are three exogenous random variables present in the system: the number of potential first and second period customers, N_1 and N_2, respectively, and the supply, Q_1. The firm knows its own supply, Q_1, and can observe the initial demand, N_1. N_1 can be thought of the customers who subscribe to a firm's email offers or who have viewed the item's page in the past. The number of potential customers in the second-period, N_2, is unknown to the firm. In contrast, customers do not (*a priori*) know the initial inventory Q_1 and the total number of customers (N_1, N_2). Formally, we let Φ denote the firm's prior belief over (Q_1, N_1, N_2), with $\Phi(q_1, n_1, n_2) = \mathbf{P}(Q_1 = q_1, N_1 = n_1, N_2 = n_2)$ denoting the joint distribution. We allow for general dependencies between the inventory level and the demand in the two periods. We slightly abuse notation and let $\Phi(q_1, n_1) = \mathbf{P}(Q_1 = q_1, N_1 = n_1)$ denote the marginal over (Q_1, N_1). (For technical reasons, we assume that N_1 has a finite support.)

Let $\Phi_C(\cdot)$ denote a customer's belief prior to receiving any information from the firm. The relationship between Φ and Φ_C is subtle: a customer, of course, knows that $N_1 \geq 1$, but her belief is not simply Φ conditioned on $N_1 \geq 1$. The following lemma[3] establishes that the customers' belief Φ_C ascribes more weight to larger values of N_1:

Lemma 3.1. *For each (q_1, n_1, n_2), we have*

$$\Phi_C(q_1, n_1, n_2) = \frac{n_1 \Phi(q_1, n_1, n_2)}{\sum_{\hat{q}_1, \hat{n}_1, \hat{n}_2} \hat{n}_1 \Phi(\hat{q}_1, \hat{n}_1, \hat{n}_2)}.$$

In particular, we have for all functions $g(\cdot)$,

$$\mathbf{E}_C[g(Q_1, N_1, N_2)] = \frac{\mathbf{E}[N_1 g(Q_1, N_1, N_2)]}{\mathbf{E}[N_1]},$$

where $\mathbf{E}[\cdot]$ and $\mathbf{E}_C[\cdot]$ represent expectations with respect to Φ and Φ_C respectively.

A customer's belief is then biased towards larger values of N_1. Intuitively, this bias in a customer's belief arises because each customer conditions on the fact that she is present in the market, an event that is more likely when the demand N_1 is larger. Such *size-bias* is common in games where the number of players is random and unknown (McAfee and McMillan 1987). A well-known example of such size-bias is the *friendship paradox*, where in a random graph of friendships, people observe a size-biased distribution of their friend's number of friends, and observe they have more friends on average than they do (Feld 1991).

The firm seeks to maximize its expected revenue by getting more customers to purchase at time 1. On the other hand, customers seek to maximize their expected utility by choosing whether to purchase at time 1 at a high price or to wait until time 2 and risk a stock-out. A customer's action depends crucially on her beliefs about (Q_1, N_1, N_2, D_1). Since the firm has more information, it seeks to persuade the customers towards buying at time 1 by sharing this information through a signaling mechanism, as we formally describe next.

[3] All proofs are provided in the appendix.

Signaling mechanisms: Let $\theta = (Q_1, N_1)$ denote the information available to the firm, and Θ denote the set of all possible values of θ. A signaling mechanism $\mathcal{S} = (S, \sigma)$ consists of two quantities. First, the mechanism fixes a set S of signals the firm may possibly send to a customer. Let $s_j \in S$ denote the signal sent by the firm to customer j, and let $\mathbf{s} = (s_1, \cdots, s_{N_1}) \in S^{N_1}$ denote the signal profile. (We assume that a customer j does not observe the signal s_k for $k \neq j$.) Then, the mechanism fixes a mapping[4] $\sigma : \Theta \times S^{\infty} \to [0, 1]$ such that for each $\theta \in \Theta$ the firm sends the signal profile $\mathbf{s} \in S^{N_1}$ with probability $\sigma(\theta, \mathbf{s})$. We let Σ denote the set of all signaling mechanisms.

An important subclass of signaling mechanisms is the class Σ_{pub} of *public* signaling mechanisms, where, rather than privately sending a signal to each customer, the firm instead announces the signal publicly to all customers. Formally, in a public signaling mechanism $\mathcal{S} = (S, \sigma)$, for each $\theta \in \Theta$ and $\mathbf{s} \in S^{N_1}$ we have $\sigma(\theta, \mathbf{s}) > 0$ only if $s_i = s_j$ for all $i, j \leq N_1$.

We assume that the firm commits to a signaling mechanism $\mathcal{S}$, which is common knowledge among the customers, prior to observing Q_1 and N_1. Formally, the model operates in the following sequence:

1. the firm publicly commits to a signaling mechanism, $\mathcal{S}$,
2. the firm observes the realizations of Q_1 and N_1 and signals to customers according to $\mathcal{S}$,
3. the N_1 first-period customers each choose to buy at price p_1 or wait, and then finally
4. the N_2 second-period customers arrive and the remaining inventory is sold to the remaining customers at price p_2.

Each such choice of a signaling mechanism then induces a game of incomplete information among customers, and the firm seeks to maximize the expected revenue in a resulting Bayesian equilibrium, which we describe formally next.

[4] We require the mapping σ to satisfy $\sigma(\theta, \mathbf{s}) = 0$ when $\mathbf{s} \notin S^{N_1}$ and $\sum_{\mathbf{s} \in S^{\infty}} \sigma(\theta, \mathbf{s}) = 1$.

Equilibrium: Given a signaling mechanism $\mathcal{S} = (S, \sigma)$, a strategy for customer i is a function $f_i : S \to [0,1]$, where $f_i(s)$ denotes the probability customer i chooses to buy at time 1 upon receiving the signal s. Let $\mathbf{f} = (f_1, \cdots, f_{N_1})$ denote a strategy profile. Each customer i's utility depends on the number $\widehat{D}_{-i}$ of other customers who elect to buy at time 1. Given $(Q_1, N_1, \widehat{D}_{-i})$, let $h(Q_1, N_1, \widehat{D}_{-i})$ denote the incremental expected utility of customer i for buying at time 1 over waiting until time 2:

$$h(Q_1, N_1, \widehat{D}_{-i}) \triangleq u_1(Q_1, \widehat{D}_{-i}) - \mathbf{E}_C \left[u_2(Q_1, N_1, N_2, \widehat{D}_{-i}) \mid Q_1, N_1, \widehat{D}_{-i} \right]$$

where the expectation is over the number N_2 of customers arriving at time 2. Given a signaling mechanism $\mathcal{S}$, a strategy profile $\mathbf{f}$ constitutes an *equilibrium* if each customer's strategy maximizes her expected utility, assuming all others follow theirs:

$$f_i(s) = \begin{cases} 1 & \text{if } \mathbf{E}_C^{\mathcal{S},\mathbf{f}}[h(Q_1, N_1, \widehat{D}_{-i}) | s_i = s] > 0; \\ 0 & \text{if } \mathbf{E}_C^{\mathcal{S},\mathbf{f}}[h(Q_1, N_1, \widehat{D}_{-i}) | s_i = s] < 0, \end{cases} \tag{3.1}$$

where $\mathbf{E}_C^{\mathcal{S},\mathbf{f}}[\cdot]$ (and analogously $\mathbf{E}^{\mathcal{S},\mathbf{f}}[\cdot]$) denotes the expectation induced by the signaling mechanism $\mathcal{S}$, the strategy profile $\mathbf{f}$, and the belief Φ_C (resp., Φ).

Firm's decision problem: The firm seeks to maximize her expected revenue by persuading customers to buy at time 1. Let $r(Q_1, N_1, D_1) = \mathbf{E}[R(Q_1, N_1, N_2, D_1) \mid Q_1, N_1, D_1]$ denote the firm's expected revenue when D_1 customers buy at time 1.

Given a signaling mechanism $\mathcal{S}$ and customer equilibrium $\mathbf{f}$, the expected revenue to the firm is given by $R(\mathcal{S}, \mathbf{f}) = \mathbf{E}^{\mathcal{S},\mathbf{f}}[r(Q_1, N_1, D_1)]$. The revenue-optimal (private) signaling mechanism is then the solution to the following optimization problem:

$$\max_{\mathcal{S} \in \Sigma} \mathbf{E}^{\mathcal{S},\mathbf{f}}[r(Q_1, N_1, D_1)] \text{ subject to } \mathbf{f} \text{ satisfying (3.1).} \tag{3.2}$$

In order to find the optimal public signaling mechanism, one approach would be to analyze a similar optimization problem, where the maximization is performed over the set of public signals Σ_{pub}. However, as we discuss below, this direct approach is challenging. Instead, we take the indirect approach of analyzing problem (3.2), characterizing the structure of its optimal solution, and establishing that the optimal mechanism is public.

Despite customers being homogeneous, it is not immediate that the optimal signaling mechanism is public. In fact, we can find settings where private signaling is necessary to generate optimal revenue with homogeneous customers. We include an example in Appendix B.3.

3.3 Formulation of the firm's decision problem

In this section, we simplify the firm's decision problem and formulate it as a linear program over a smaller decision space. As a step towards that formulation, we first show that the firm can focus on a subclass of signaling mechanisms. In the following, when there is no ambiguity, we drop the subscript from Q_1, N_1, and D_1.

The following lemma, which follows from a revelation-principle style argument (Bergemann and Morris 2018), states that for the problem of finding the optimal signaling mechanism, the firm can restrict its attention to *symmetric* signaling mechanisms that use binary signals, denoting recommendations to buy at time 1 ($s_i = 1$) or to wait until time 2 ($s_i = 0$), and for which the equilibrium action for each customer i is to follow the recommendation. Using standard terminology, we refer to such signaling mechanisms as symmetric *direct mechanisms*.

Lemma 3.2. *Consider any signaling mechanism $\mathcal{S} = (S, \sigma)$ with equilibrium $\mathbf{f}$. There exists a private signaling mechanism $\mathcal{U} = (U, \upsilon)$ with equilibrium $\mathbf{g}$ such that*

1. *$U = \{0, 1\}$, with $g_i(0) = 0$ and $g_i(1) = 1$ for each i;*
2. *υ is* symmetric*: $\upsilon(Q, N, \boldsymbol{s}) = \upsilon(Q, N, \boldsymbol{s}')$ whenever $\boldsymbol{s}'$ is a permutation of $\boldsymbol{s}$.*
3. *The firm's revenue under $(\mathcal{U}, g)$ is equal to that under $(\mathcal{S}, f)$.*

We emphasize that the equivalent mechanism $\mathcal{U}$ in Lemma 3.2 need not be public, even if the mechanism $\mathcal{S}$ is. In particular, this implies that for the firm's problem of finding the optimal public mechanism, the reduction to direct mechanisms does not hold. It is for this reason that the problem of finding optimal public signaling mechanisms is, in general, difficult (Dughmi and Xu 2017). Nevertheless, by directly analyzing the decision problem (3.2), we show in Section 3.4 that the optimal signaling mechanism is indeed public.

Note that in such an equilibrium, due to the underlying symmetry, the number $\widehat{D}_{-i}$ of customers other than customer i who choose to buy at time 1 has the same distribution for each i. Henceforth, we drop the subscript on $\widehat{D}_{-i}$ and let $\widehat{D}$ denote the number of other customers buying at time 1 from a fixed customers' perspective. This symmetry presents a further simplification: the firm's problem reduces to identifying, for each (Q, N), the number of customers $D \in \{0, \cdots, N\}$ to recommend buying now. Subsequently, the firm selects a subset of D customers uniformly at random, and sends them the signal $s = 1$, whereas the rest receive the signal $s = 0$. Abusing the notation slightly, we let $\sigma(Q, N, D)$ denote the probability that, given inventory Q and demand N, the signaling mechanism σ recommends D customers to buy now.

The preceding simplifications allow us to represent the firm's problem (3.2) as a linear program. Formally, let $\pi(q, n, d) = \Phi(q, n)\sigma(q, n, d)$ denote the (prior) joint probability that the firm has inventory $Q = q$, with demand $N = n$ and the firm recommends $D_1 = d$ customers to buy now. Then, using Lemma 3.1 and the fact that for any customer i we have $\mathbf{P}(s_i = 1|Q, N, D) = D/N$ for all (Q, N, D), we obtain the following linear programming formulation in π of (3.2).

$$\max_{\pi} \sum_{(q,n)\in\Theta} \sum_{d=0}^{n} r(q,n,d)\cdot\pi(q,n,d)$$

$$\text{subject to,} \quad \sum_{(q,n)\in\Theta} \sum_{d=1}^{n} d\cdot h(q,n,d-1)\cdot\pi(q,n,d) \geq 0, \tag{3.3a}$$

$$\sum_{(q,n)\in\Theta} \sum_{d=0}^{n-1} (n-d)\cdot h(q,n,d)\cdot\pi(q,n,d) \leq 0, \tag{3.3b}$$

$$\sum_{d=0}^{n} \pi(q,n,d) = \Phi(q,n), \quad \text{for all } (q,n)\in\Theta, \tag{3.3c}$$

$$\pi(q,n,d) \geq 0, \quad \text{for all } d\in\{0,\cdots,n\} \text{ and } (q,n)\in\Theta. \tag{3.3d}$$

Here, the constraints (3.3a) and (3.3b) capture the two cases in the definition (3.1). A feasible solution π corresponds to the symmetric direct signaling mechanism σ given by $\sigma(q,n,d) = \pi(q,n,d)/\Phi(q,n)$. A detailed derivation of (3.3) is provided in Appendix B.2.

3.4 Optimality of public signaling

Observe that if for each q,n we have $\sigma(q,n,d) = 0$ for all $d \notin \{0,n\}$, then the direct mechanism σ always recommends either all customers to buy or all of them to wait. Thus, this mechanism can be implemented publicly by announcing the recommendation publicly to all customers. (As mentioned earlier, the converse does not hold: not all public mechanisms can be represented this way.) Our main result of this section shows that the optimal solution to (3.3) satisfies this condition and hence is a public mechanism:

Theorem 3.1. *There exists an optimal signaling mechanism σ that is public: $\sigma(q,n,0) + \sigma(q,n,n) = 1$ for all q,n. In addition, this public signaling mechanism has $\sigma(q,n,n) = 1$ for all q,n such that $h(q,n,n-1) \geq 0$.*

This theorem is theorem is proven in two steps:

1. First, we show that any optimal mechanism will tell all customers to buy for all states (Q, N) where customers would buy under full-information.
2. Then, we show that any mechanism where $D \notin \{0, N\}$ customers are asked to buy with non-zero probability can be manipulated to weakly increase revenue.

This approach rests on two critical lemmas that describe the nature of the competition among the customers. In fact, Theorem 3.1 applies for any instance of (3.3), assuming r is monotonically increasing in d and h satisfies these lemmas. To prove the first step of the theorem, we make use of this first lemma, which states that if a customer prefers to buy now when some number of other customers buy now, then she continues to prefer buying now if all other customers buy now.

Lemma 3.3. *For any (q, n), if there exists a $d \in \{1, \cdots, n-1\}$ such that $h(q, n, d-1) \geq 0$, then $h(q, n, n-1) \geq 0$.*

We note that the condition in Lemma 3.3 is weaker than *strategic complementarity* (Morris and Shin 2003): instead of having a customer's utility for taking an action increase as more customers take that action, the customer's preferred action of buying now remains unchanged if all customers choose that action. Also, this condition is weaker than the notion of *single crossing* (Morris and Shin 2003), since we do not require a customer's optimal action to remain unchanged as more customers buy now, just that it remains optimal to buy now when all customers do.

In a symmetric signaling mechanism, the higher the probability that a customer is asked to buy now, the more likely it is that other customers are *also* asked to buy now, implying more competition. This then induces a trade-off for the customer: does a customer prefer lower competition upon receiving a recommendation to buy, but with a higher likelihood of not receiving such recommendation, or does she prefer higher likelihood of receiving a recommendation to buy, albeit with more competition? In order to prove the second step of the theorem, we use the following lemma, which specifies how customers make this trade-off.

Lemma 3.4. *The incremental utility function h satisfies the following properties:*

(3.4.1) *For any (q, n) with $h(q, n, n-1) \geq 0$, we have*

$$\frac{d}{n} \cdot h(q, n, d-1) \leq h(q, n, n-1).$$

(3.4.2) *For any (q, n) with $h(q, n, n-1) < 0$, we have*

$$\frac{d}{n} \cdot h(q, n, d-1) \leq \left(\frac{r(q, n, d) - r(q, n, 0)}{r(q, n, n) - r(q, n, 0)} \right) \cdot h(q, n, n-1).$$

Note that for $(Q, N) = (q, n)$ with $h(q, n, n-1) \geq 0$, each customer would prefer to buy now if all other customers were buying now. In such a setting, property (3.4.1) states that a customer receives higher expected utility when all customers are asked to buy now than when a random d of them are asked to buy now. Thus, customers prefer higher likelihood of being asked to buy now, even if that induces more competition. For values of $(Q, N) = (q, n)$ where $h(q, n, n-1) < 0$, customers prefer waiting until time 2, even if all other customers buy now. In this setting, even though property (3.4.1) no longer holds, a weaker property (3.4.2) holds.

The proof of Theorem 3.1 uses Lemma 3.3 and Lemma 3.4 to show that any feasible solution to (3.3) can be altered to one with a public structure (i.e., $\pi(q, n, d) = 0$ for $d \notin \{0, n\}$) without decreasing the objective. More specifically, the first step of the proof shows that $\sigma(q, n, n) = 1$ when $h(q, n, n-1) \geq 0$, and the second step shows that any solution where $\sigma(q, n, d) > 0$ for $d \notin \{0, n\}$ can be weakly improved to another where $\sigma(q, n, d) = 0$. Full details are given in Appendix B.1.

This structural form of the optimal solution also leads to a simple algorithm that computes the optimal mechanism in $O(|\Theta|)$ time, assuming $r(q, n, n)$ and $h(q, n, n-1)$ for $(q, n) \in \Theta$

are already computed or can be computed in time $O(|\Theta|)$. This algorithm is obtained by reformulating the linear program as a fractional knapsack problem, as we describe below.

Let S be the set of pairs (q, n) such that $h(q, n, n-1) < 0$. Theorem 3.1 implies $\sigma(q, n, n) = 1$ for all $(q, n) \in \Theta \setminus S$ and $\sigma(q, n, n) \in [0, 1]$ for all $(q, n) \in S$. This guarantees (3.3b) holds and allows us to rewrite (3.3a) as

$$\sum_{(q,n)\in S} n|h(q, n, n-1)| \cdot \Phi(q, n) \cdot \sigma(q, n, n) \quad \leq \sum_{(q,n)\in\Theta\setminus S} nh(q, n, n-1) \cdot \Phi(q, n).$$

Thus, (3.3a) can be seen as a fractional knapsack constraint, where the knapsack capacity is $W = \sum_{(q,n)\in\Theta\setminus S} nh(q, n, n-1)\Phi(q, n)$, the items are the pairs $(q, n) \in S$, and the weight and value of the item (q, n) are respectively given by $w_{q,n} = n|h(q, n, n-1)|\Phi(q, n)$ and $v_{q,n} = r(q, n, n)$, respectively. The amount of each item (q, n) in the optimal solution to this fractional knapsack problem gives the value of $\sigma(q, n, n)$ in the optimal signaling mechanism.

Given the capacity W, the weights $w_{q,n}$, and the values $v_{q,n}$, the optimal solution to the fractional knapsack problem can be found in time linear in the number of items (Korte and Vygen 2012). Since this algorithm requires finding the values $(q, n) \in \Theta \setminus S$ and computing W, which takes $O(|\Theta|)$ time, the optimal signaling mechanism can be found in time $O(|\Theta|)$.

Finally, under additional assumptions on the distribution of the demand and the inventory, Theorem 3.2 allows us to obtain further structural results on the optimal public mechanism. In particular, when the demand in the two periods are conditionally independent given the inventory, the optimal signaling mechanism has a threshold structure, as defined below.

Theorem 3.2. *Suppose the demands N and N_2 are conditionally independent, given the inventory Q. Then, there exists an optimal signaling mechanism, σ, that has a threshold structure: for each q there exists an $N_q \in \mathbb{N}$ and $x_q \in (0, 1]$ such that $\sigma(q, n, n) = 1$ if $n > N_q$, $\sigma(q, n, 0) = 1$ if $n < N_q$ and $\sigma(q, n, n) = x_q$ and $\sigma(q, n, 0) = 1 - x_q$ if $n = N_q$.*

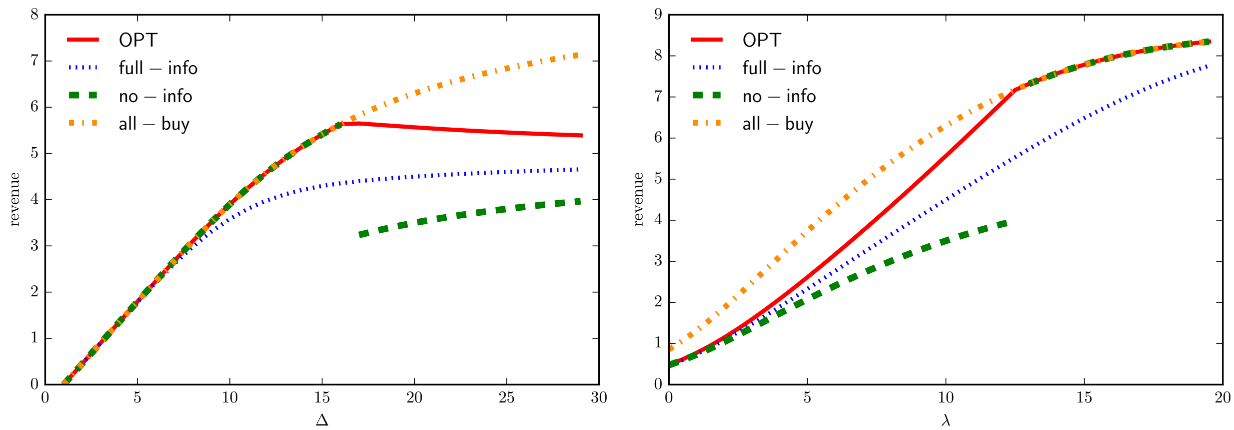
The proof is given in Appendix B.1.

3.4.1 Revenue comparison

To support and quantify our analytical results, we numerically compare the revenue generated by the optimal mechanism against three benchmarks. The first benchmark ("$\mathsf{no-info}$") is the signaling mechanism that provides no information to the customer; all customers act based solely on their beliefs Φ_b. The second benchmark ("$\mathsf{full-info}$") is the signaling mechanism that reveals to the customers all the relevant information, namely the inventory level Q and number of customers N. Finally, as a third benchmark ("$\mathsf{all-buy}$"), we consider the revenue when all N customers buy at time 1, irrespective of whether it is in their best interest to do so. Unlike the first two benchmarks, the final benchmark may not be achievable through a signaling mechanism. Despite this, it provides a useful metric to evaluate the revenue increase through signaling.

For our numerical computations, we assume that there is no new demand at time 2, i.e., $N_2 = 0$. Furthermore, we assume that the inventory level Q and the number of customer N are independent: $\Phi(q,n) = \mathbf{P}(Q = q, N = n) = P_\Delta(q) \cdot P_\lambda(n)$ where P_Δ denotes the uniform distribution over $\{0, 1, \cdots, \Delta - 1\}$ and P_λ denotes the Poisson distribution with mean $\lambda > 0$. Note that for a fixed λ, as Δ increases, the inventory level is higher, and consequently customers are more likely to wait until time 2 to buy. Similarly, for a fixed Δ as λ increases, the demand increases and the customers are more likely to buy at time 1. For our analysis, we let $v = 1, p_1 = 0.9, p_2 = 0.5$, and $c = 0.1$. We begin with a brief discussion of the no-information and the full-information benchmarks.

In the no-information setting, there may be multiple equilibria, depending on the parameters. For high inventory (large Δ) compared to demand, the only equilibrium will be for all customers to wait to purchase the product. For low inventory, the only equilibrium will be for all customers to buy now. For moderate inventory, both may be equilibria, along with an equilibria where customers randomize their decision. When there are multiple options, we



(a) Revenue for $\lambda = 10$.

(b) Revenue for $\Delta = 20$.

Figure 3.1: Comparing signaling mechanisms.

plot the revenue of the revenue-maximizing equilibria. In Figure 3.1a, the no-information revenue has a discontinuity precisely when all customers buying now stops being an equilibria.

Under full-information benchmark, (Q, N) is revealed to all the customers. Note that, just as in the no-information mechanism, there may be multiple equilibria under the full-information mechanism. For example, there will be at least two equilibria whenever $h(Q, N, N-1) \geq 0$ but $h(Q, N, 0) < 0$, one where all customers buy at time $t-1$ and another where all buy at time 2. Hereafter, to obtain meaningful comparisons, we focus on the equilibria with the highest expected revenue for the firm. As in the no-information setting, for low values of Δ, the full-information mechanism produces near-optimal revenue. For higher values of Δ, the revenue approaches 5, which corresponds to the case where all customers wait until time 2 to buy at price $p_2 = 0.5$.

In Figure 3.1, we compare the revenue achieved by our three mechanisms[5]. When the inventory level is low relative to the demand, we see that the optimal signaling mechanism is to provide no information; all customers already prefer to buy now, and signaling any information can only induce customers to wait and thus reduce revenue. On the other hand,

[5]While the full-information and no-information mechanisms may possess multiple equilibria, we focus on the equilibrium with the highest revenue.

Table 3.1: Efficacy of public signaling.

λ	Δ	$R_{\mathsf{no-info}}$	$R_{\mathsf{full-info}}$	R_{OPT}	$R_{\mathsf{all-buy}}$	$M_{\mathsf{no-info}}$	$M_{\mathsf{full-info}}$
5	10	1.85	2.32	2.79	3.33	63%	46%
5	20	2.20	2.45	2.71	3.96	28%	16%
5	30	2.31	2.48	2.64	4.15	18%	9%
5	50	2.39	2.49	2.59	4.30	10%	5%
10	20	3.74	4.73	5.73	6.74	66%	49%
10	30	4.17	4.83	5.50	7.50	39%	24%
10	50	4.50	4.90	5.30	8.10	22%	12%
20	30	6.56	9.14	11.71	11.81	98%	96%
20	50	8.00	9.60	11.20	14.40	49%	33%
30	50	10.50	14.09	17.69	18.89	85%	74%

when the inventory level is moderate or high relative to the demand, we observe that the optimal public signaling mechanism achieves substantially higher expected revenue than the no-information and full-information mechanisms.

To measure the efficacy of signaling, in Table 3.1 we compare the revenue achieved by the optimal signaling mechanism against the "$\mathsf{all-buy}$" benchmark. To make this comparison, we introduce two metrics, $M_{\mathsf{no-info}}$ and $M_{\mathsf{full-info}}$. Let $R_{\mathsf{no-info}}, R_{\mathsf{full-info}}$, and R_{OPT} denote respectively the expected revenue under the no-information, the full-information, and the optimal mechanisms, and let $R_{\mathsf{all-buy}}$ denote that in the "$\mathsf{all-buy}$" setting. We define the metrics $M_{\mathsf{no-info}} \triangleq \frac{R_{\mathsf{OPT}}-R_{\mathsf{no-info}}}{R_{\mathsf{all-buy}}-R_{\mathsf{no-info}}}$ and $M_{\mathsf{full-info}} \triangleq \frac{R_{\mathsf{OPT}}-R_{\mathsf{full-info}}}{R_{\mathsf{all-buy}}-R_{\mathsf{full-info}}}$. Note that from the definition, we have $R_{\mathsf{OPT}} = M_i R_{\mathsf{all-buy}} + (1-M_i)R_i$ for $i \in \{\mathsf{no-info}, \mathsf{full-info}\}$. Thus, M_i measures how much of the difference between $R_{\mathsf{all-buy}}$ and R_i can be captured through signaling. Note that we have only included values where $R_{\mathsf{no-info}} \neq R_{\mathsf{all-buy}}$; if $R_{\mathsf{no-info}} = R_{\mathsf{all-buy}}$, all customers will buy under no-information and optimal signaling cannot increase revenue. Hence we will focus on situations where $R_{\mathsf{no-info}} < R_{\mathsf{all-buy}}$.

From Table 3.1, we observe that signaling can lead to values of $M_{\mathsf{full-info}}$ as high as 96%. Similarly, the values of $M_{\mathsf{no-info}}$ can be as high as 98%. This suggests that optimal signaling can be effective in yielding revenues close to $R_{\mathsf{all-buy}}$. We also observe that signaling is most effective in raising revenue over full or no information mechanism when the maximum inventory Δ is not too large relative to the mean demand λ, i.e., when the customers actually face a substantial stock-out risk.

3.5 Heterogeneous Customers

We briefly consider the efficacy of signaling in the setting of customer heterogeneity. Specifically, we consider the setting where there are two types of customer: high-valued and low-valued, who value the product at values v_H and v_L, respectively. We assume that both v_H and v_L are larger than p_1 so that all customers could be persuaded to buy in the first period. Pricing is agnostic to the types of customers; this is natural when sales are done on a public website but signaling can be done privately (for example, over email). We begin by assuming the firm can view each customers' type accurately and commits to a signaling mechanism that may depend on the types of customers. After, we consider the case where the firm only observes the total number of customers present in the first period. The question remains: how should the firm signal to optimize her revenue?

3.5.1 Observed types

Formally, we introduce a linear program similar to (3.3) that optimizes revenue over signaling mechanisms in this heterogeneous setting. We must first establish variables comparable to those in (3.3): Let N_H and N_L denote the number of high- and low-valued customers in the system, and let Φ denote the joint probability over Q, N_H, N_L. We let D_H and D_L denote the number of high- and low-valued customers, respectively, who buy at time 1 and let h_H and h_L denote their incremental expected utility when they buy at time 1. We let $N = N_H + N_L$,

$D = D_H + D_L$ and $\hat{D} = \hat{D}_H + \hat{D}_L$. More concretely, for each customer type i,

$$h_i(Q, N, \hat{D}) = \min\left\{\frac{Q}{\hat{D}+1}, 1\right\}(v_i - p_1) - \min\left\{\frac{(Q-\hat{D})^+}{N-\hat{D}}, 1\right\}(v_i - p_2) + c.$$

With this, the linear program can be expressed as

$$\begin{aligned}
\max_{\pi} \quad & \sum_{q,n_H,n_L,d_H,d_L} r(q, n_H + n_L, d_H + d_L)\pi(q, n_H, n_L, d_H, d_L) \\
\text{subject to,} \quad & \sum_{q,n_H,n_L,d_H,d_L} d_H h_H(q, n_H + n_L, d_H + d_L - 1)\pi(q, n_H, n_L, d_H, d_L) \geq 0 \\
& \sum_{q,n_H,n_L,d_H,d_L} d_L h_L(q, n_H + n_L, d_H + d_L - 1)\pi(q, n_H, n_L, d_H, d_L) \geq 0 \\
& \sum_{q,n_H,n_L,d_H,d_L} (n_H - d_H)\, h_H(q, n_H + n_L, d_H + d_L)\pi(q, n_H, n_L, d_H, d_L) \leq 0 \\
& \sum_{q,n_H,n_L,d_H,d_L} (n_L - d_L)\, h_L(q, n_H + n_L, d_H + d_L)\pi(q, n_H, n_L, d_H, d_L) \leq 0 \\
& \sum_{d_H,d_L} \pi(q, n_H, n_L, d_H, d_L) = \Phi(q, n_H, n_L) \\
& \pi(q, n_H, n_L, d_H, d_L) \geq 0, \quad \text{for all } q, n_H, n_L, d_H, d_L.
\end{aligned}$$

Each feasible solution to this LP corresponds to a private signaling mechanism σ, where the probability of asking d_H high-type and d_L low-type customers to buy, when the state is (q, n_H, n_L), is given by $\sigma(q, n_H, n_L, d_H, d_L) = \frac{\pi(q,n_H,n_L,d_H,d_L)}{\Phi(q,n_H,n_L)}$. We denote the optimal revenue achievable by a private signaling mechanism by R_{pri}.

Although private signaling is the most general form of signaling mechanisms, such private signaling may be difficult to logistically implement. Because of this, we now consider three subclasses of signaling mechanisms, which are more restrictive than private signals, but may be more practically desirable. In each of these classes, we continue to restrict the firm to signal pure actions (or action profiles) and only consider those mechanisms under which obedience is an equilibrium. To highlight these restrictions, we call such signals *recommendations*.

The least restrictive class among the three is the class of *type-based recommendations*, where the same recommendation is privately sent to all customers of the same type, but customers do not observe the signals sent to customers of a different type. Such mechanisms arise, for example, when targeted emails are sent to different email lists consisting of different customer types. Note that type-based recommendations are still private mechanisms, since the signals are not announced publicly to all customers. On the other hand, *public recommendations* announce publicly which customer types should buy and which types should wait. Specifically, under public recommendations, each customer is aware of the actions recommended to all customer types. Such mechanisms arise when all communication reaches both types of customer, like when this information is communicated on a public-facing website. Most restrictively, *blind public recommendations* require the firm to recommend the same action to all customer types. This could occur when the firm is limited to a "low stock" signal on its website. We let the expected revenue from the optimal type-based recommendation, public recommendation, and blind public recommendation mechanisms be denoted by $R_{\mathsf{type}}, R_{\mathsf{pub}}$ and R_{blind}, respectively. Through a simple argument, the optimal revenues can be shown to satisfy the following ordering.

Proposition 1. *The optimal revenues satisfy* $R_{\mathsf{pri}} \geq R_{\mathsf{type}} \geq R_{\mathsf{pub}} \geq R_{\mathsf{blind}}$.

Proof of Proposition 1. Any type-based recommendation mechanism can be viewed as a private mechanism, so $R_{\mathsf{pri}} \geq R_{\mathsf{type}}$. Similarly, every public blind recommendation can be viewed as a public recommendation where it is never the case that one type of customers is suggested an action that another type is not: hence, $R_{\mathsf{pub}} \geq R_{\mathsf{blind}}$. Finally, consider any public recommendation mechanism. Since the equilibrium is obedient, a customer type follows her recommendation irrespective of what the other customer type is recommended. This implies that the customer type would follow her recommendation even if she is not informed of the other customer types' recommendation. Hence, a public recommendation mechanism and the obedient equilibrium can be implemented as a type-based recommendation mechanism, with the same revenue, yielding $R_{\mathsf{type}} \geq R_{\mathsf{pub}}$. □

3.5.2 Unobserved types

We now consider the setting where the firm cannot observe the types of individual customers. Instead, the firm must signal according to their beliefs about the number of each customer type. Let N denote the total number of customers arrived who appear in the first period. For simplicity, we let each given customer have a high-type independently and with probability γ.[6] The firm then must send a signal to each customer without knowledge of their type.

By the same argument as in the proof of Lemma 3.2, customers need not play a mixed strategy in equilibrium. Further, given any signaling mechanism $\mathcal{S} = (S, \sigma)$, there exists a symmetric signaling mechanism $\mathcal{U} = (U, v)$ that achieves the same revenue such that $U = \{0, 1, 2\}$ where high-type customers play strategy $g_H(u) = \mathbf{I}\{u \geq 1\}$ and low-type customers play strategy $\mathbf{I}\{u \geq 2\}$. This follows from a very similar argument to that in Lemma 3.2.

With this, we can, once more, construct a linear program. The decision variables will be of the form $\sigma(Q, N, D_2, D_1)$, which gives the probability of sending D_2 customers the signal 2 and D_1 customers the signal 1. Unlike in previous settings, given a choice of D_2 and D_1, it is unclear how many customers will attempt to buy in the first period: the customers who receive D_1 will buy if and only if they are of high-type. Hence, in the expression of the linear program, we let $D_H \leq D_1$ denote the number of customers sent the signal D_1 who are of

[6] The distribution of total number of high-type customers given N is a binomial with parameters N and γ

high-type and choose to buy now.

$$\max_{\pi} \sum_{q,n,d_2,d_1} \pi(q,n,d_2,d_1) \sum_{d_H=0}^{d_1} \binom{d_1}{d_H} \gamma^{d_H}(1-\gamma)^{d_1-d_H} r(q,n,d_2+d_H)$$

$$\text{subject to,} \quad \sum_{q,n,d_2,d_1} \pi(q,n,d_2,d_1) \sum_{d_H=0}^{d_1} \binom{d_1}{d_H} \gamma^{d_H}(1-\gamma)^{d_1-d_H} d_1 h_H(q,n,d_2+d_H-1) \geq 0$$

$$\sum_{q,n,d_2,d_1} \pi(q,n,d_2,d_1) \sum_{d_H=0}^{d_1} \binom{d_1}{d_H} \gamma^{d_H}(1-\gamma)^{d_1-d_H} d_2 h_L(q,n,d_2+d_H-1) \geq 0$$

$$\sum_{q,n,d_2,d_1} \pi(q,n,d_2,d_1) \sum_{d_H=0}^{d_1} \binom{d_1}{d_H} \gamma^{d_H}(1-\gamma)^{d_1-d_H} d_1 h_L(q,n,d_2+d_H-1) \leq 0$$

$$\sum_{q,n,d_2,d_1} \pi(q,n,d_2,d_1) \sum_{d_H=0}^{d_1} \binom{d_1}{d_H} \gamma^{d_H}(1-\gamma)^{d_1-d_H} (n-d_2-d_1) h_H(q,n,d_2+d_H-1) \leq 0$$

$$\sum_{d_2,d_1} \pi(q,n,d_2,d_1) = \Phi(q,n)$$

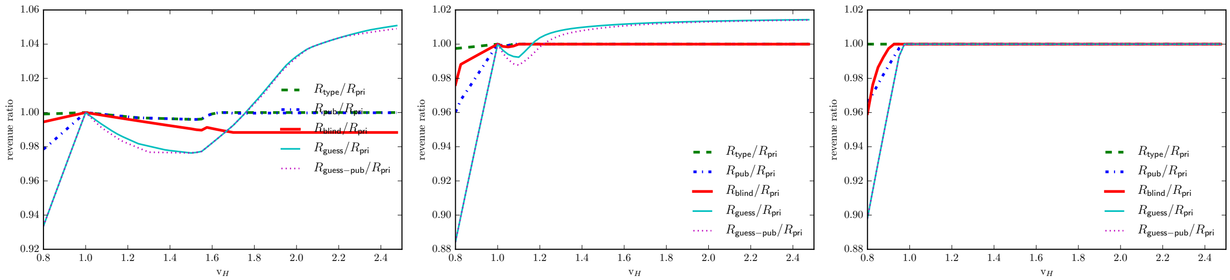
$$\pi(q,n,d_2,d_1) \geq 0, \quad \text{for all } q,n,d_2,d_1.$$

Like before, each feasible solution to this LP corresponds to a private signaling mechanism σ, where the probability of asking d_2 customers to buy now and asking d_1 customers to buy only if they are high-type, when the state is (q,n), is given by $\sigma(q,n,d_2,d_1) = \frac{\pi(q,n,d_2,d_1)}{\Phi(q,n)}$. We denote the optimal revenue achievable by a private mechanism by R_{unobs}. We will also consider the optimal revenue when the firm may only send the same recommendation to all customers in the below, and denote it $R_{\mathsf{unobs-pub}}$.

3.5.3 Numeric comparison

Note that, when $v_H = v_L$, our results for the homogeneous customers types, specifically Theorem 3.1, imply that all the inequalities above are equalities. Below, through numerical computations, we seek to understand the performance of these mechanisms when $v_H \neq v_L$.

In Figure 3.2, we plot the ratio of $R_{\mathsf{type}}, R_{\mathsf{pub}}$ R_{blind}, R_{unobs} and $R_{\mathsf{unobs-pub}}$ to R_{pri} to see how closely to private signaling each subclass performs. We focus first on Figure 3.2a, where



(a) $\Delta = 5$, $v_L = 1$, $p_1 = 0.8$, $p_2 = 0.5$, and $N = 2, 3, 4$.

Figure 3.2: Revenue ratios.

we fix $\Delta = 5$, $v_L = 1$, $p_1 = 0.8$, $p_2 = 0.5$, $c = 0$, and let the number of total customers be distributed uniformly over $\{1, 2, 3, \ldots, N\}$, with each customer being high-type with probability probability $\frac{1}{2}$. With these variables defined, we vary v_H (note that we allow values where $v_H < 1 = v_L$).

As Figure 3.2a shows, when $v_H \neq v_L$, private signaling performs better than each of the other classes of signaling mechanisms. In particular, unlike the case of $v_H = v_L$, public recommendations do not generate the optimal revenue. Nor is it true that the optimal signaling mechanism recommends the same action to all customers of a type: the optimal signaling mechanism achieves strictly better revenue than the optimal type-based recommendation mechanism. In practical terms, the firm cannot generate the optimal revenue with even a conservative definition of a "public" mechanism: the firm must recommend different actions to customers of the same type. The poor performance of R_{unobs} and $R_{\mathsf{unobs-pub}}$ show that observing a customer's type provides a significant revenue advantage.

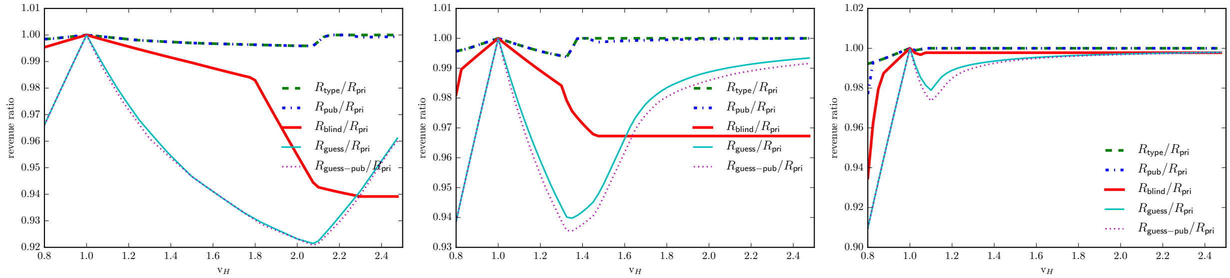
However, we observe that the both the type-based and the public recommendation mechanism perform reasonably well for most values of v_H. In fact, for large values of v_H, both of these recommendations achieves revenue close to the optimal (private) signaling mechanisms. This could arise from type H customers becoming increasingly indifferent upon hearing that only they (and not type L customers) are asked to buy. Conversely, for high values of v_H, blind public recommendations do poorly because the requirement that recommendations are incentive compatible for both types to follow becomes very restrictive. In fact, blind public mechanisms are outperformed by unobserving mechanisms for large v_H. Note that all three of these classes of mechanisms perform closer to optimal for larger values of N.

Beyond these plots, we find that for typical parameters, public recommendations generates 85-95% of the revenue of private signaling. In fact, in our numeric search, the worst ratio we could find between public recommendations and private signaling was 66%. These results suggest that even under customer heterogeneity, public recommendations are effective in raising the firm's revenue.

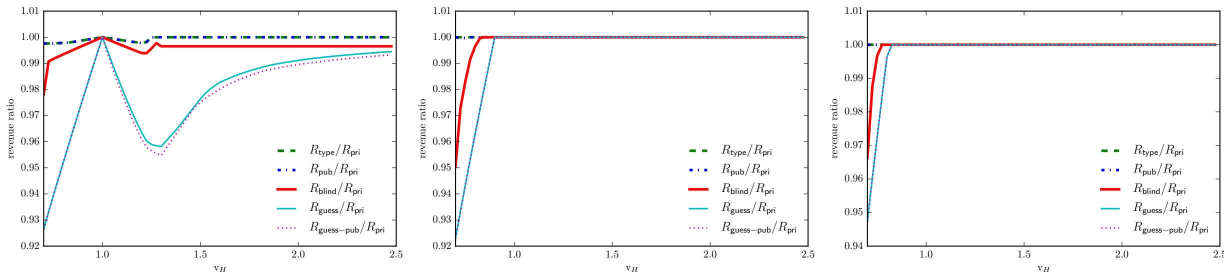
In each of Figures 3.2b, 3.2c, 3.2d and 3.2e, we modify the parameters used in Figure 3.2a: In Figure 3.2b, we let $\Delta = 7$. In Figures 3.2c and 3.2d, we let p_1 equal 0.7 and 0.9, respectively. Finally, in Figure 3.2e, we let the number of customer (N) be uniform over $\{1, \ldots, N\}$ and let $\gamma \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$, instead of restricting $\gamma = \frac{1}{2}$.

3.6 Discussion

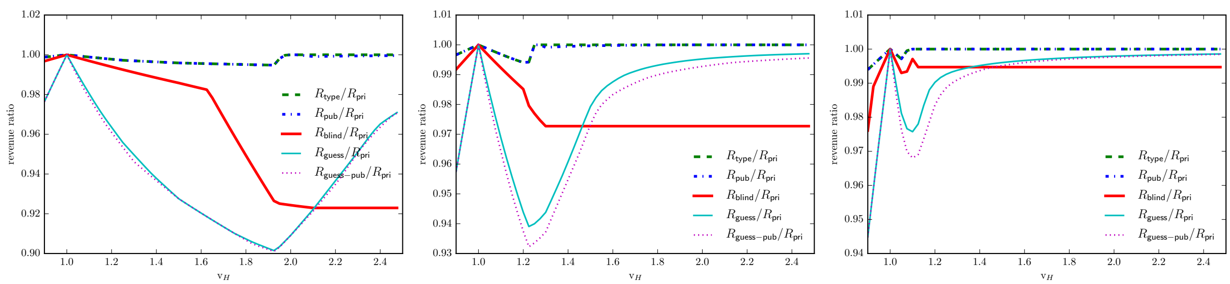
We consider inventory signaling with commitment and establish that the optimal mechanism is public. Moreover, we show that with commitment the firm achieves significantly higher revenue over no information sharing, unlike the setting with cheap talk. Our analytical result relies on the assumption that customers are homogeneous. When customers are heterogeneous, simple public direct mechanisms are no longer optimal in general: it is not hard to construct examples where private signals outperform public direct mechanisms. However, our numerical investigations show that such mechanisms continue to perform close to optimal when customer types are sufficiently differentiated. When there are heterogeneous customers, we find private signaling outperforms public signaling, but public signaling performs close to optimal when



(b) $\Delta = 7$, $v_L = 1$, $p_1 = 0.8$, $p_2 = 0.5$, and $N = 2, 3, 4$.



(c) $\Delta = 5$, $v_L = 1$, $p_1 = 0.7$, $p_2 = 0.5$, and $N = 2, 3, 4$.



(d) $\Delta = 5$, $v_L = 1$, $p_1 = 0.9$, $p_2 = 0.5$, and $N = 2, 3, 4$.

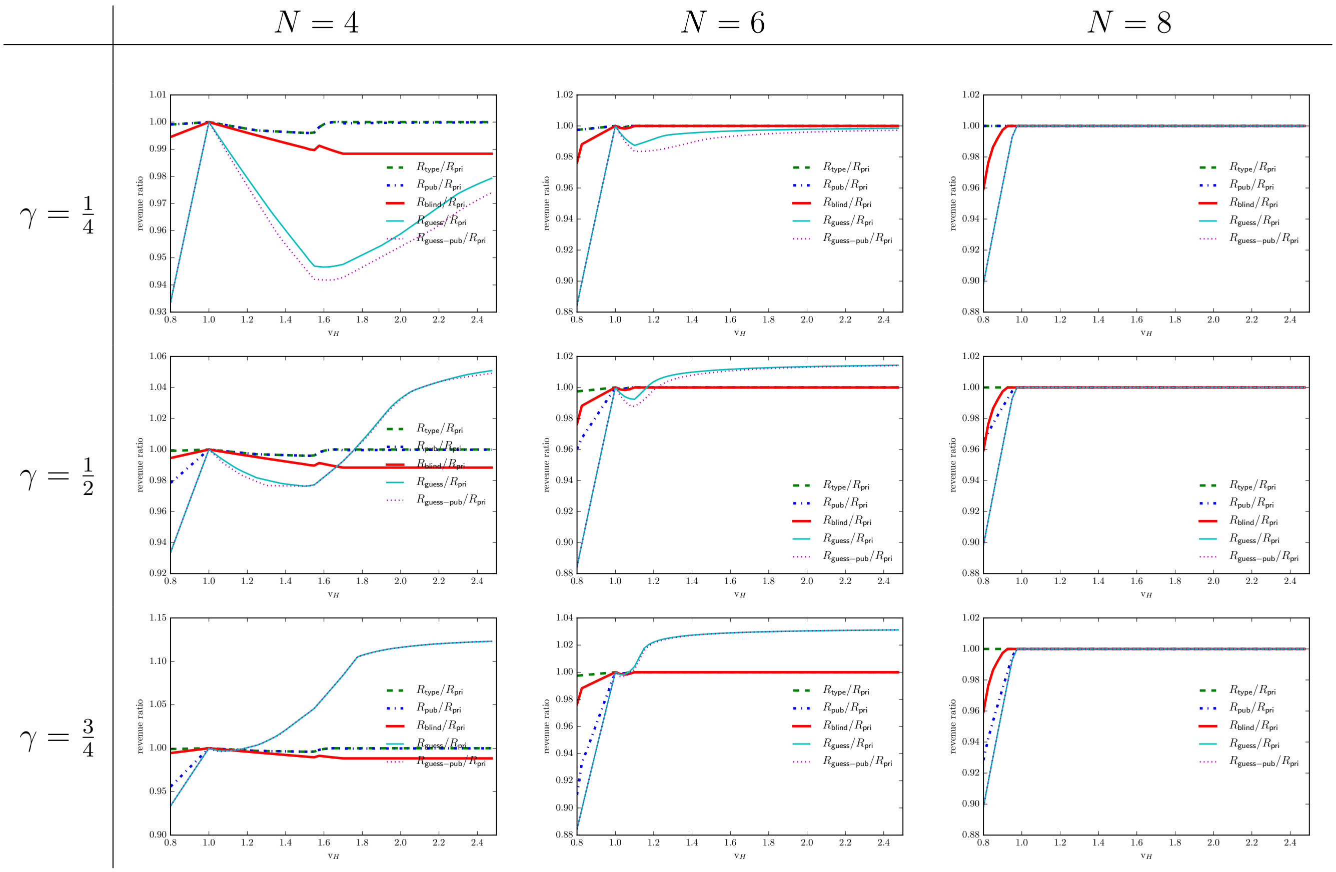
Figure 3.2: Revenue ratios.

customer types are sufficiently differentiated.

Our model assumes that the price in the second period is fixed. However, the firm may wish to set a lower price when remaining quantity is higher, to further encourage customers to buy. One might instead consider a setting where the price in the second period is a function of the remaining supply after time 1, i.e., $p_2 \triangleq p_2(Q - D)$. In the case where $p_2(\cdot)$ is non-increasing, convex, log-concave and $N_2 = 0$, our results can be extended.

Theorem 3.3. *If $N_2 = 0$ and the second round price $p_2(\cdot)$ is:*

- *decreasing:* $p_2(n) \geq p_2(n + 1)$,



(e) $\Delta = 5$, $v_L = 1$, $p_1 = 0.9$, $p_2 = 0.5$, and $H, L \in \{2, 3, 4\}$.

Figure 3.2: Revenue ratios.

- *convex:* $\Delta p_2(n) \leq \Delta p_2(n+1)$,
- *and concave:* $\frac{\Delta p_2(n)}{p_2(n)} \geq \frac{\Delta p_2(n+1)}{p_2(n+1)}$,

the optimal signaling mechanism is public.

It remains an open question whether the optimal signaling mechanism is public in the case of a more complex pricing function $p_2(\cdot)$ or a non-zero demand N_2.

Our model further assumes that the quantity, Q, is an observed exogenous random variable. This models the settings like Fulfilled by Amazon, where Amazon does not control the supply, but simply signals inventory based on it. We can extend our results to settings where the firm has control over the supply, and can set it strategically.

Theorem 3.4. *Suppose N is independent of Q, and the firm can choose the distribution of Q.*

(3.4.1) *If the firm chooses a distribution for Q prior to observing N_1, the optimal signaling mechanism is public.*

(3.4.2) *If the firm chooses a distribution for Q after observing N_1, the optimal signaling mechanism is public.*

In our analysis, the firm is careful to treat the customers as Bayesian and to let their beliefs be a posterior formed by the firm's prior updated with the event that they are present in the system. When the firm does not do this, and assumes customers share their prior instead of having a size-biased version of it, a couple undesirable outcomes may occur, which are now described. Suppose customers each have a payoff-irrelevant feature that is communicated to (or even generated by) the firm. The firm could then choose the optimal manner of signaling with a linear program much like the one given in Section 3.5, but with non size-biased incentive compatibility constraints[7] and with $v_H = v_L$. First, signaling according to this errant description of customer beliefs may lead to a mechanism that is not incentive-compatible: while the "buy now" signal leads to zero customer surplus for the firm's imagined customer, a real customer's would receive negative surplus because their prior puts more weight on larger values of N. Hence, not all customers would follow their signals, since doing so would not be an equilibrium. Further, we find examples where the optimal signaling mechanism signals differently for each type, which raises issues of fairness and equality.

The reader should note that while we limit our focus to signaling in online retail, our results and methods apply far more broadly. Indeed, consider a symmetric game with many agents, action set $\{0, 1\}$ and where the principal's utility is monotonically increasing in the number of customers who choose action 1. If customers' incremental utility satisfies

[7] Which each divide the inequalities by N_H or N_L, depending on the constraint

Lemmas 3.3 and 3.4, then Theorem 3.1 applies, and the optimal signaling mechanism is public.

We now proceed to Chapter 4, where we consider customers who are not expected utility maximizers.

CHAPTER 4

PERSUADING RISK-CONSCIOUS AGENTS: A GEOMETRIC APPROACH

If no one ever took risks,
Michaelangelo would have painted
the Sistine floor.

Neil Simon

4.1 Introduction

Given the inherent informational asymmetries in various online marketplaces between the platform's operators and its users, information design can potentially play a important role in marketplace design. Starting from the seminal papers of Rayo and Segal (2010) and Kamenica and Gentzkow (2011), the methodology of Bayesian persuasion and information design has received a lot of recent attention from the academic community. A number of papers have applied Bayesian persuasion to various application contexts, such as engagement/misinformation in social networks (Candogan and Drakopoulos 2017), queueing systems (Lingenbrink and Iyer 2018a), and online retail (Lingenbrink and Iyer 2018b, Drakopoulos et al. 2018).

In most of the previous work, the standard assumption is that the agent being persuaded is an expected utility maximizer. Although this assumption is well-supported from a theoretical perspective via axiomatic characterizations (Machina 1995), it is conventionally accepted that human behavior is not adequately explained by the central tenets of the theory. In particular, there is a long line of work in theoretical economics literature that studies systematic biases in human behavior that have been well-documented empirically (Kahneman and Tversky 1972, Tversky and Kahneman 1992, Rabin 1998, Genesove 2001). Owing to this disparity between human behavior and the expected utility model, the methodology of Bayesian persuasion may not satisfactorily apply to information design problems in real marketplaces and other practical settings.

Motivated by this concern, our main goal in this chapter is to extend the methodology of Bayesian persuasion to settings where the receiver may not be an expected utility maximizer. Specifically, we allow for general models of the receiver's utility under uncertainty, where the receiver's utility is a general function of their beliefs. We term such receivers as *risk-conscious*.[1] The only assumption we make on the receiver's utility is that it is continuous in the receiver's belief. With this assumption, we study the sender's persuasion problem of optimally sharing information about payoff-relevant uncertainty with the receiver to affect the latter's actions.

When agents are expected utility maximizers, a revelation-principle style argument is often invoked to reduce the set of possible messages (i.e., *signals*) the sender might send to the set of actions available to the receiver. This reduction simplifies the sender's optimization problem substantially, and the resulting problem can be written as a linear program with one *obedience* constraint corresponding to each receiver action. In contrast, with risk-conscious agents, due to the non-linearity of the receiver's utility, we note that the revelation-principle style argument fails to hold. Because of this, one is forced to consider signaling schemes where the set of signals could be as large as the set of all beliefs for the agent. In turn, this renders the usual approach to finding the optimal signaling scheme ineffective.

The main contribution of this chapter is in overcoming this technical challenge by adopting a more direct geometric approach. In particular, for general persuasion settings, we show that the sender's problem can be reduced to solving a *convex* program in variables that denote the joint probability of the underlying state and the receiver's actions, and whose optimal solution can be decomposed to obtain the optimal signaling scheme. Using this characterization, and

[1] We emphasize that this is different from risk-aversion modeled as expectation of a concave function. Any such utility function must necessarily be linear in the beliefs.

results from convex analysis, we further show that the number of signals in the optimal signaling scheme is at most the product of the number of states and the number of actions available to the receiver.

We formulate a convex program with domain the convex hull of the set of beliefs where each fixed action is optimal for the receiver. For general persuasion settings, this set might be quite complex. To understand the implications of our model, we study a specific persuasion setting, namely *binary persuasion*, where the receiver's actions are binary $a \in \{0, 1\}$, and the sender always prefers that the receiver choose action 1 over 0. Under a mild convexity assumption on the receiver's utility function, we show that the convex optimization problem obtained in fact reduces to a linear program whose solution can be efficiently computed. Furthermore, by analyzing this linear program, we obtain a canonical construction to identify the set of signals that the optimal signaling scheme adopts in any binary persuasion setting. This canonical set of signals involves a combination of pure signals, which fully reveal the underlying state, and binary signals, which induce in the receiver uncertainty between two pure states. Finally, we show that in contrast to expected utility maximizing receivers, a hesitant risk-conscious receiver can sometimes be *fully persuaded* to choose $a = 1$, even if the optimal action is $a = 0$ under her prior beliefs.

To illustrate our methodology, we consider persuasion in a setting with endogenous priors. In particular, we analyze the model of a queueing system introduced in Lingenbrink and Iyer (2018a), where the service provider seeks to persuade arriving customers to join an unobservable single server queue to wait for service. Inspired by the literature on the psychology of waiting in queues (Maister et al. 1984), we extend their model to allow for customers with general risk-conscious utility. We show that our theoretical results carry over to this setting, despite the state space being infinite and the endogeneity of the prior. We investigate the special case where the customer has a mean-stdev utility (Nikolova 2010, Nikolova and Stier-Moses 2014, Cominetti and Torrico 2016), where her utility for joining the queue is the sum of her mean waiting time and a multiple of its standard deviation. We numerically find that the optimal signaling scheme has an interesting "sandwich" structure that induces an ordering among the various "join" signals (i.e., signals for which the customer's optimal action is to wait for service) in terms of the variance of the resulting waiting times.

4.1.1 Literature Review

Our results contribute to the study of *Bayesian persuasion* (Kamenica and Gentzkow 2011, Rayo and Segal 2010, Bergemann and Morris 2018, 2016a, Taneva 2019, Doval and Ely 2016). Our work particularly takes influence from Kamenica and Gentzkow (2011), where the sender's choice of an optimal signaling scheme is considered geometrically: the optimal signaling scheme can be viewed as a Bayes-plausible distribution of beliefs. Additionally, our work uses a reduction to a linear program similar to that done in Kolotilin et al. (2016). For a review on the general methods of Bayesian persuasion and information design, see Kamenica (2018). For a methodological approach to Bayesian persuasion in finite settings, see Bergemann and Morris (2018) and Taneva (2019). Bayesian persuasion has been applied to many operational settings, such as where the sender seeks to persuade the receiver to engage in an online platforms (Candogan and Drakopoulos 2017), prepare for a potential disaster (Alizamir et al. 2018), join a queue (Lingenbrink and Iyer 2018a), fairly price-discriminate (Bergemann et al. 2015), make a purchase with limited inventory (Lingenbrink and Iyer 2018b, Drakopoulos et al. 2018), or participate in a matching platform (Romanyuk and Smolin 2018), in addition to many other settings.

In Bayesian persuasion, the receiver is generally treated as an expected utility maximizer and has some function over states of which she is maximizing the expectation. This does not cover cases where the receiver's utility function cannot be expressed as an expectation of the state. Machina (1995) outlined the limitation of this treatment, and showed that these models assume a “linearity in the probabilities.” In our work, we consider a risk-conscious customer who may neither be risk-neutral or have this linearity in probabilities.

This form of utility function is closely related to the notion of *risk measure* in financial mathematics (Föllmer and Schied 2011). In finance, apart from expected return, an investor might also care about measures that are non-linear functions of the distribution such as standard deviation of return (Markowitz 1952), value at risk (Jorion 2006), and expected shortfall (Carlo Acerbi 2002). Similarly, our risk-conscious agent’s utility can involve non-linear functions of the distribution such as variability of payoff. Nevertheless, there are differences. To be a good measure of financial risk, a risk measure in finance should be *coherent*, that is, it should satisfy a number of properties that make sense in the context of portfolio management. Many coherent risk measures (Wang 1996, Ahmadi-Javid 2012) do not have an interpretation as natural descriptions of human behavior. Conversely, and perhaps more importantly, utilities do not have to be coherent to empirically capture aspects of human decision-making. Prospect theory (Kahneman and Tversky 1979, Tversky and Kahneman 1992), the most widely used model in behavioral economics, is not coherent.

Finally, in this chapter we apply our results to a specific Bayesian persuasion problem: signaling in an $M/M/1$ queue with risk-conscious customers. Maister et al. (1984) suggested people do not treat wait-time in a queue in a linear fashion, and hence it is a setting our results have practical importance. We adapt the model from Lingenbrink and Iyer (2018a) and use the mean-standard deviation risk measure used in Nikolova and Stier-Moses (2015) to benchmark our results against the no-information and full-information signaling schemes. Note that in this setting, the state space is infinite, unlike our other results. And, as Gentzkow and Kamenica (2016) suggested, our standard approach is difficult with an infinite state space, since it may involve calculating a concave closure on the infinite-dimensional space of distributions over the state space. However, our results do hold in this setting.

4.2 Model

In the following, we present the model of Bayesian persuasion with risk-conscious agents. Our development of the model follows closely to that of the standard Bayesian persuasion setting Kamenica and Gentzkow (2011), Kamenica (2018).

4.2.1 Setup

We consider a persuasion problem with one *sender* and one *receiver*. Let X be a payoff-relevant random variable with support on a known set $\mathcal{X}$. We assume that neither the receiver nor the sender observes X. However, as we describe below, the sender has more information about X than the receiver, and seeks to use this information to influence the receiver's actions.

Formally, we assume that the distribution of X depends on the state of the world $\bar{\omega}$, which is in the finite set Ω, and is observed by the sender but not the receiver. We denote the distribution of X, conditional on $\bar{\omega} = \omega$, by F_ω. The distributions $\{F_\omega : \omega \in \Omega\}$ are commonly known between the sender and the receiver, and both share a common prior $\mu^* \in \Delta(\Omega)$[2] about the state of the world $\bar{\omega}$. For each $\mu \in \Delta(\Omega)$, we let F_μ be the distribution of X if $\bar{\omega}$ is distributed as μ, and let X_μ denote a variable distributed as F_μ. It follows immediately that $F_\mu = \sum_\omega \mu(\omega) F_\omega$.

As in the standard Bayesian persuasion setting, we assume that the receiver is Bayesian and that the sender can commit to a *signaling scheme* to affect receiver's choice of an action (which is described below in detail). A signaling scheme (S, π) consists of a signal space S and a joint distribution $\pi \in \Delta(\Omega \times S)$ such that the marginal of π over Ω equals μ^*: for

[2]Throughout, for any set S, we let $\Delta(S)$ denote the set of probability distributions over S. When S is finite, we consider $\Delta(S)$ as a subset of $\mathbb{R}^{|S|}$, endowed with the Euclidean topology.

each $\omega \in \Omega$, $\pi(\omega, S) = \mu^*(\omega)$. Specifically, we assume that, if the realized state is $\bar{\omega} = \omega$, the sender draws a signal $\bar{s} \in S$ according to the conditional distribution $\pi(\cdot|\bar{\omega} = \omega)$, and conveys it to the receiver. For simplicity of notation, we denote a signaling scheme (S, π) by the joint distribution π, and let $\pi_\omega \in \Delta(S)$ denote the conditional distribution of $\bar{s}$ given $\bar{\omega} = \omega$. Throughout, we assume that the sender's commits to a signaling scheme prior to observing the state $\bar{\omega}$, and that the sender's choice π of the signaling scheme is common knowledge among the sender and the receiver.

As mentioned above, we assume that the receiver is Bayesian. Given the sender's signaling scheme π, upon observing the signal $\bar{s} = s$, the receiver updates her belief about the state using Bayes' rule (whenever possible) from her prior μ_0 to posterior $\mu_s(\cdot) \in \Delta(\Omega)$. In particular, we have for all $\omega \in \Omega$,

$$\mu_s(\omega) = \frac{\pi(\omega, s)}{\sum_{\omega' \in \Omega} \pi(\omega', s)}, \tag{4.1}$$

whenever the denominator on the right-hand side is positive.[3] Note that this implies that upon receiving the signal $\bar{s} = s$, the receiver believes that the payoff-relevant variable X is distributed as F_{μ_s}.

4.2.2 Actions, strategy and utility

Upon observing the signal $\bar{s}$, the receiver chooses an action $a \in A$, where the set of actions A is assumed to be finite. Given a signaling scheme π, the receiver's strategy $a(\cdot)$ specifies an action[4] $a(s)$ for each realization $s \in S$ of the signal $\bar{s}$.

We let $v(\omega, a)$ denote the sender's utility in state $\bar{\omega} = \omega$ when the receiver chooses the action $a \in A$. Furthermore, we assume that the sender is an expected utility maximizer.

[3] We let $\mu_s \in \Delta(\Omega)$ be arbitrary if the denominator is zero.

[4] Although our definition implies a pure strategy, we can easily incorporate mixed strategies where the receiver chooses an action at random. We suppress this technicality for the sake of readability.

(One can represent the utility function $v(\omega, a)$ as an expectation of a utility function $\hat{v}(X, a)$ over the payoff-relevant variable X and the action a, conditional on $\bar{\omega} = \omega$; we suppress the details for brevity.)

Our point of departure from the standard framework is in the definition of the receiver's utility. Specifically, we relax the assumption that the receiver is an expected utility maximizer; as we describe next, our setup allows for more general models of receiver's utility over the uncertain outcome X. We refer to such receivers as being *risk-conscious*.

Formally, for any belief $\mu \in \Delta$ of the receiver, we assume that the receiver's *utility* upon taking an action $a \in A$ is given by $\hat{\rho}(F_\mu, a) \in \mathbb{R}$. For notational simplicity, we define the *utility function* $\rho : \Delta(\Omega) \times A \to \mathbb{R}$ as $\rho(\mu, a) \triangleq \hat{\rho}(F_\mu, a)$. Given a belief μ, we assume that the receiver chooses an action $a \in A$ that achieves the highest utility $\rho(\mu, a)$.

Note that a receiver is an expected utility maximizer if and only if, for each $a \in A$, the utility function $\rho(\mu, a)$ is linear[5] in μ. In particular, there exists a function $u : \mathcal{X} \times \Omega \times A \to \mathbb{R}$ such that $\rho(\mu, a) = \mathbf{E}\left[u(X, \bar{\omega}, a) | \bar{\omega} \sim \mu\right]$ for all $\mu \in \Delta(\Omega)$ and $a \in A$, if and only if the receiver is an expected utility maximizer (with vNM utility function u). Our setup therefore includes as a special case the standard Bayesian persuasion framework with an expected utility maximizing receiver. However, the generality of our setting allows us to capture a much wider range of receiver behavior.

As an illustration, consider the setting of a customer deciding whether or not to wait for service in a queue. The receiver's utility depends on her unknown wait time X, and suppose the queue-operator observes some correlated feature ω (congestion, server availability, etc). A natural model Nikolova (2010), Nikolova and Stier-Moses (2014), Cominetti and Torrico (2016) for a risk-conscious customer posits that the customer only waits for service if,

[5]Note that if $\rho(\mu, a)$ is a utility function of an agent, then so is $g(\rho(\mu, a))$ for any increasing function $g : \mathbb{R} \to \mathbb{R}$. Thus, this linearity holds only up to an increasing transformation. We suppress such (irrelevant) transformations for the sake of clarity.

given her beliefs, the mean of her waiting time plus a multiple of its standard deviation is below a threshold. (Such a behavioural model may arise from the customer's requirement for service reliability, or from her desire to plan her day subsequent to service completion.) This model can be captured in our setting by letting $A = \{0, 1\}$ and assuming, for example, that $\rho(\mu, 1) = \tau - \left(\mathbf{E}[X_\mu] + \beta \sqrt{\mathrm{Var}(X_\mu)}\right)$ and $\rho(\mu, 0) = 0$. It is straightforward to check that $\rho(\mu, 1)$ is not linear in μ.

Throughout this chapter, we make the following assumption on ρ:

Assumption 4.1. *For each $a \in A$, the utility function $\rho(\cdot, a)$ is continuous.*[6]

4.2.3 Persuasion game

We are now ready to describe the sender's persuasion problem. First, we require that for any choice of the signaling scheme π, the receiver's strategy maximizes her utility: for each $s \in S$, we have

$$a(s) \in \arg\max_{a \in A} \rho(\mu_s, a). \tag{4.2}$$

We call any strategy that satisfies (4.2) an optimal strategy for the receiver. Given an optimal strategy $a(\cdot)$, the sender's expected utility for choosing a signaling scheme π is given by[7]

$$\mathbf{E}[v(\bar{\omega}, a(\bar{s}))|(\bar{\omega}, \bar{s}) \sim \pi].$$

The sender seeks to choose a signaling scheme π that maximizes her expected utility, assuming that the receiver responds with an optimal strategy.[8] Thus, the sender's persuasion problem can be posed as

[6]Here, continuity is with respect to the Euclidean topology on $\Delta(\Omega)$.

[7]We use the notation $\mathbf{E}[f(Y)|Y \sim \mu]$ for the expectation of $f(Y)$ where Y is distributed as μ.

[8]When the receiver has multiple optimal strategies, we assume that the sender chooses her most preferred one; the literature refers to this as the sender-preferred subgame-perfect equilibrium Kamenica and Gentzkow (2011) with a strategy satisfying (4.2).

$$\begin{aligned} &\max_{\pi \in \Delta(\Omega \times S)} \mathbf{E}[v(\bar{\omega}, a(\bar{s})) | (\bar{\omega}, \bar{s}) \sim \pi] \\ \text{subject to,} \quad & a(s) \in \arg\max_{a \in A} \rho(\mu_s, a), \quad \text{for all } s \in S, \\ & \pi(\omega, S) = \mu^*(\omega), \quad \text{for all } \omega \in \Omega. \end{aligned} \tag{4.3}$$

Our main goal in this chapter is to find and characterize the sender's optimal signaling scheme to the persuasion problem (4.3). Note that this problem, as posed, is challenging, as it requires first choosing an optimal set of signals S, and a joint distribution π over $\Omega \times S$. Without an explicit bound on the cardinality of set S, the persuasion problem seems intractable. In the next section, we reframe the problem to obtain a tractable formulation.

4.3 Towards a tractable formulation

In Bayesian persuasion literature, the receiver is treated as an expected utility maximizer, and a revelation-principle style argument is frequently used to restrict attention to signaling schemes to ones where the signal space S satisfies $|S| = |A|$. Before we discuss our approach for general risk-conscious agents, we provide a more detailed discussion of this argument, and discuss why it fails in our setting.

The revelation-principle style argument rests on the following observation: when the receiver is an expected utility maximizer, if two signals s_1 and s_2 both lead to the same optimal action $a(s_1) = a(s_2) = a$, then a is still an optimal action for the receiver if the signaling scheme reveals only that $\bar{s} \in \{s_1, s_2\}$ whenever it was supposed to reveal s_1 or s_2. This property is straightforward to show by using the linearity of the utility function $\rho(\mu, a)$ for expected utility maximizers. One can then use this property to coalesce all signals that lead to the same optimal action for the receiver into an *action recommendation.* Such a coalesced signaling scheme has at most $|A|$ signals, and moreover, the agent's optimal strategy is *obedient*, i.e., it is optimal for the agent to follow the action recommendation.

However, when the receiver is risk-conscious, the preceding argument may no longer hold. This is because, when signals with same optimal action are coalesced, the resulting posterior of the receiver on the coalesced signal changes, and without linearity, the receiver's optimal action may change (Nikolova 2010). Since the receiver's actions may change, it no longer suffices to only consider signaling schemes with action recommendations.

Despite this difficulty, a version of the preceding observation, which we term *coalescence*, continues to hold with a risk-conscious receiver. To see this, observe that if two signals $\bar{s} = s_1$ and $\bar{s} = s_2$ lead to the same posterior $\mu \in \Delta(\Omega)$ for the receiver, then the receiver's posterior is still μ if the signaling scheme reveals only that $\bar{s} \in \{s_1, s_2\}$ whenever it was supposed to reveal s_1 or s_2. This coalescence property follows immediately from the fact that the receiver's posterior beliefs are expectations and expectations are linear. Thus, using the same argument as before, the coalescence property allows us to coalesce all signals that lead to the same posterior belief of the receiver into a *belief recommendation.* In such a coalesced signaling scheme, we can take the signal space S to be $\Delta(\Omega)$, the set of posteriors. Furthermore, in such a scheme, if the receiver is recommended a belief μ, her posterior belief is indeed μ.

Summarizing the preceding discussion, we can write the sender's persuasion problem (4.3) as

$$\begin{aligned} &\max_{\pi \in \Delta(\Omega \times \Delta(\Omega))} \mathbf{E}[v(\bar{\omega}, a(\bar{s}))|(\bar{\omega}, \bar{s}) \sim \pi] \\ &\text{subject to,} \quad a(s) \in \arg\max_{a \in A} \rho(\mu_s, a), \quad \text{for all } s \in \Delta(\Omega), \\ &\qquad \pi(\omega, \Delta(\Omega)) = \mu^*(\omega), \quad \text{for each } \omega \in \Omega, \\ &\qquad \mu_s = s, \quad \text{for all } s \in \Delta(\Omega). \end{aligned} \tag{4.4}$$

Although we have characterized the set of signals, this is still a challenging problem because of the size of the set $\Delta(\Omega \times \Delta(\Omega))$. To make further progress, we use the notion of Bayes-plausibility introduced by Kamenica and Gentzkow (2011). We state the following result, without proof, from Kamenica and Gentzkow (2011):

Lemma 4.1 (Bayes-plausibility Kamenica and Gentzkow (2011))**.** *A signaling scheme* $\pi \in$

$\Delta(\Omega \times \Delta(\Omega))$ *satisfies the condition* $\mu_s = s$ *for almost all* $s \in \Delta(\Omega)$*, only if the measure* $\eta(\cdot) = \pi(\Omega, \cdot) \in \Delta(\Delta(\Omega))$ *is Bayes-plausible, i.e., only if the following condition holds:*

$$\mathbf{E}[\bar{s}(\omega)|\bar{s} \sim \eta] = \mu^*(\omega), \quad \textit{for each } \omega \in \Omega.$$

Conversely, for any Bayes-plausible measure η*, the signaling scheme defined as* $\pi(\omega, ds) = s(\omega)\eta(ds)$ *satisfies* $\mu_s = s$ *for all* $s \in \Delta(\Omega)$*.*

The preceding lemma allows us to optimize over the space of Bayes-plausible measures $\eta \in \Delta(\Delta(\Omega))$. These are probability measures over the set of posteriors $\Delta(\Omega)$ with the property that the mean of the distribution of the induced posteriors equals the prior. Furthemore, observe that for any Bayes-plausible measure η, the sender's expected utility under the corresponding signaling scheme $\pi(\omega, ds) = s(\omega)\eta(ds)$ can be written as

$$\begin{aligned}
\mathbf{E}[v(\bar{\omega}, a(\bar{s}))|(\bar{\omega}, \bar{s}) \sim \pi] &= \mathbf{E}\left[\mathbf{E}\left[v(\bar{\omega}, a(\bar{s}))|\bar{\omega} \sim \bar{s}\right] \;\middle|\; \bar{s} \sim \eta\right] \\
&= \mathbf{E}\left[\sum_{\omega \in \Omega} \bar{s}(\omega) v(\omega, a(\bar{s})) \;\middle|\; \bar{s} \sim \eta\right].
\end{aligned}$$

Thus the sender's expected utility can be written as a function of the receiver's strategy and the probability measure η. Taken together, we obtain an intermediate reformulation of the sender's problem:

$$\begin{aligned}
&\max_{\eta \in \Delta(\Delta(\Omega))} \sum_{\omega \in \Omega} \mathbf{E}\left[\bar{s}(\omega) v(\omega, a(\bar{s})) \;\middle|\; \bar{s} \sim \eta\right] \\
&\text{subject to,} \quad a(s) \in \arg\max_{a \in A} \rho(\mu_s, a), \\
&\qquad\qquad \mathbf{E}[\bar{s}(\omega)|\bar{s} \sim \eta] = \mu^*(\omega), \quad \text{for each } \omega \in \Omega.
\end{aligned} \tag{4.5}$$

To state our first result, we need a definition and some notation. First, define the following sets:

$$\mathcal{P}_a \triangleq \left\{\mu \in \Delta(\Omega) : a \in \arg\max_{a' \in A} \rho(\mu, a)\right\}, \quad \text{for each } a \in A.$$

The set $\mathcal{P}_a$ denotes the set of posteriors for which the action $a \in A$ is optimal for the receiver. Note that, by the continuity of $\rho(\cdot, a)$, the set $\mathcal{P}_a$ is closed (and hence, compact) for each

$a \in A$. Morever, we have $\cup_{a\in A}\mathcal{P}_a = \Delta(\Omega)$ and if $\mu \in \mathcal{P}_a \cup \mathcal{P}_{a'}$, then both actions a and a' are optimal for the receiver with posterior μ. Next, we need some notation: for any set $A \in \mathbb{R}^m$, let $\mathsf{Conv}(A)$ denote the convex hull of A, defined as:

$$\mathsf{Conv}(A) = \left\{ y : y = \sum_{i=1}^{j} \lambda_i x_i, \text{for some } j \geq 1,\ \lambda_i \geq 0, x_i \in A \text{ for all } 1 \leq i \leq j \text{ and } \sum_{i=1}^{j} \lambda_i = 1 \right\}$$

Then, we have the following lemma, which essentially states that corresponding to each Bayes' plausible measure, there exists a real b_a and a vector m_a for each action a such that the sender's utility under η can be written as a bi-linear function of m_a and b_a. Thus, the lemma allows us to directly optimize over m_a and b_a, instead of over Bayes-plausible measures η.

Lemma 4.2. *For any Bayes-plausible measure $\eta \in \Delta(\Delta(\Omega))$ and optimal receiver strategy $a(\cdot)$, there exists $\{(b_a, m_a)\}_{a\in A}$, with $b_a \geq 0$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ for each $a \in A$, such that*

$$\sum_{a\in A} b_a m_a = \mu^*, \tag{4.6}$$

$$\mathbf{E}\left[\bar{s}(\omega)v(\omega, a(\bar{s}))\middle|\bar{s} \sim \eta\right] = \sum_{a\in A} b_a m_a(\omega) v(\omega, a) \quad \textit{for each } \omega \in \Omega. \tag{4.7}$$

Conversely, for any $\{(b_a, m_a) : b_a \geq 0$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ for each $a \in A\}$ satisfying (4.6), *there exists a Bayes-plausible measure η and an optimal receiver strategy $a(\cdot)$ such that* (4.7) *holds.*

Before proving the lemma, we provide some interpretation for the quantities m_a and b_a. For each a, the quantity b_a denote the probability that the receiver plays action a under the optimal strategy, when the sender uses the Bayes-plausible measure η. Similarly, m_a denotes the distribution of the state $\bar{\omega}$, conditioned on the receiver choosing action a. Note that m_a may not correspond to any actual posterior that the receiver holds about the state. However, as we show in the following proof, each m_a corresponds to the mean of all posteriors that the receiver holds conditioned on her playing action a. Figure 4.1 gives some geometric intuition for these quantities, as well as the distributions η_0 and η_1 introduced in the following proof.

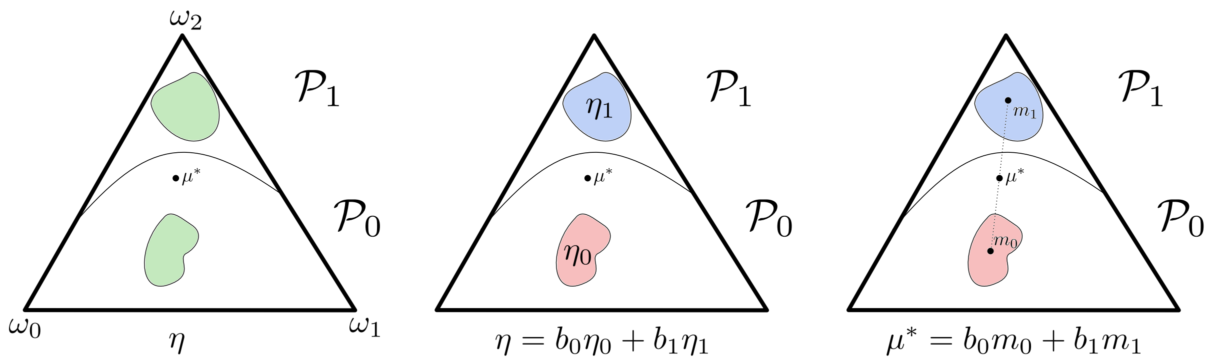


Figure 4.1: In the first figure, $\eta \in \Delta(\Delta(\Omega))$ has support in the green region. Since η is Bayes-plausible, $\mathbf{E}[\bar{s}(\omega)|\bar{s} \sim \eta] = \mu^*(\omega)$ for each $\omega \in \Omega$. In the second figure, we separate η into η_0 and η_1, where η_a has support in $\mathcal{P}_a$. In the third figure, we depict m_0 and m_1, where $\mathbf{E}[\bar{s}(\omega)|\bar{s} \sim \eta_a] = m_a^*(\omega)$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$.

Proof. Fix any Bayes-plausible measure η and an optimal receiver strategy $a(\cdot)$. For each $a \in A$, define $b_a \triangleq \mathbf{P}(a(\bar{s}) = a|\bar{s} \sim \eta) \in [0, 1]$ to be the probability that the receiver chooses action a under the corresponding signaling scheme. For each $a \in A$, if $b_a = 0$, then let η_a be any probability measure with support on $\mathcal{P}_a$. Otherwise, define η_a to the measure obtained by conditioning η on the event $\bar{s} = a$. More precisely, we have $\eta_a(ds) \triangleq \frac{1}{b_a}\mathbf{I}\{a(s) = a\}\eta(ds)$ if $b_a > 0$. Note that, by the definition of the sets $\mathcal{P}_a$, the support of η_a is $\mathcal{P}_a$ for each $a \in A$. The following expressions are immediate:

$$\sum_{a \in A} b_a \eta_a = \eta, \quad \sum_{a \in A} b_a = 1.$$

We let $m_a(\omega) \triangleq \mathbf{E}[\bar{s}(\omega)|\bar{s} \sim \eta_a]$ for each $\omega \in \Omega$. In words, $m_a(\omega)$ is the *mean posterior belief* of the receiver that the state is ω, given that she chooses the action $a(s) = a$. From the Bayes-plausibility of η, we obtain for each $\omega \in \Omega$:

$$\sum_{a \in A} b_a m_a = \sum_{a \in A} b_a \, \mathbf{E}[\bar{s}(\omega)|\bar{s} \sim \eta_a] = \mathbf{E}[\bar{s}(\omega)|\bar{s} \sim \eta] = \mu^*,$$

where the second equality follows from the fact that $\sum_{a \in A} b_a \eta_a = \eta$. Finally, it is straightforward to verify that $m_a \in \mathsf{Conv}(\mathcal{P}_a)$, since m_a is the mean of the posterior distribution η_a

with support on $\mathcal{P}_a$, and $\mathcal{P}_a$ is closed and compact. Finally, note that for each $\omega \in \Omega$,

$$\begin{aligned}\sum_{a\in A} b_a m_a(\omega) v(\omega, a) &= \sum_{a\in A} b_a\, \mathbf{E}[\bar{s}(\omega)|\bar{s} \sim \eta_a] v(\omega, a) \\ &= \sum_{a\in A} b_a\, \mathbf{E}[\bar{s}(\omega) v(\omega, a(\bar{s}))|\bar{s} \sim \eta_a] \\ &= \mathbf{E}\left[\bar{s}(\omega) v(\omega, a(\bar{s}))\middle|\bar{s} \sim \eta\right],\end{aligned}$$

where the first equality follows from the fact that $a(\bar{s}) = a$ when $\bar{s} \sim \eta_a$, and the second equality follows from the fact that $\sum_{a\in A} b_a \eta_a = \eta$.

Conversely, suppose we have $\{(b_a, m_a)\}_{a\in A}$ with $b_a \in [0, 1]$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ with $\sum_{a\in A} b_a m_a = \mu^*$. By the definition of the convex hull, $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ implies the existence of $\{(\mu_i^a, \lambda_i^a) : i = 1, \cdots, j_a\}$ such that $\mu_i^a \in \mathcal{P}_a$ and $\lambda_i^a \geq 0$ for each $i \leq j_a$ with $\sum_{i=1}^{j_a} \lambda_i^a = 1$ and $m_a = \sum_{i=1}^{j_a} \lambda_i^a \mu_i^a$. Define $\eta \in \Delta(\Delta(\Omega))$ to be the discrete distribution that selects the posterior μ_i^a with probability $b_a \lambda_i^a$. Then, we have for all $\omega \in \Omega$,

$$\mathbf{E}[\bar{\mu}(\omega)|\bar{\mu} \sim \eta] = \sum_{a\in A}\sum_{i=1}^{j_a} b_a \lambda_i^a \mu_i^a(\omega) = \sum_{a\in A} b_a \left(\sum_{i=1}^{j_a} \lambda_i^a \mu_i^a(\omega)\right) = \sum_{a\in A} b_a m_a(\omega) = \mu^*(\omega).$$

This proves the Bayes-plausibility of η. Finally, define the strategy $a(\cdot) : \Delta(\Omega) \to A$ so that $a(\mu_i^a) = a$ for each $i \leq j_a$ and $a \in A$, and for other values of μ, let $a(\mu)$ be an arbitrary element in $\arg\max_{a\in A} \rho(\mu, a)$. Since $\mu_i^a \in \mathcal{P}_a$, it is straightforward to verify that the strategy $a(\cdot)$ is optimal. Finally, we have

$$\begin{aligned}\sum_{a\in A} b_a m_a(\omega) v(\omega, a) &= \sum_{a\in A} b_a \sum_{i=1}^{j_a} \lambda_i^a \mu_i^a(\omega) v(\omega, a) \\ &= \sum_{a\in A}\sum_{i=1}^{j_a} (b_a \lambda_i^a) \cdot \mu_i^a(\omega) \cdot v(\omega, a(\mu_i^a)) \\ &= \sum_{a\in A}\sum_{i=1}^{j_a} \eta(\mu_i^a) \cdot \mu_i^a(\omega) \cdot v(\omega, a(\mu_i^a)) \\ &= \mathbf{E}\left[\bar{s}(\omega) v(\omega.\bar{s})|\bar{s} \sim \eta\right].\end{aligned}$$

Here, the first equation follows from the fact that $m_a = \sum_{i=1}^{j_a} \mu_i^a \lambda_i^a$, the second equation follows from the fact that $a(\mu_i^a) = a$, and the third equation follows from the definition of η. This completes the proof of the lemma. $\square$

Using the preceding lemma, we can now reframe the sender's persuasion problem as

$$\begin{aligned}
&\max_{\{b_a, m_a : a \in A\}} \sum_{\omega \in \Omega} \sum_{a \in A} b_a m_a(\omega) v(\omega, a) \\
&\text{subject to, } \sum_{a \in A} b_a m_a = \mu^*, \\
&\qquad m_a \in \mathsf{Conv}(\mathcal{P}_a), \quad b_a \in [0,1] \quad \text{for each } a \in A.
\end{aligned} \tag{4.8}$$

Note that the preceding problem is extremely simple when compared to (4.5); the optimization is over $|A|(1+|\Omega|)$ real variables, and the variables m_a belong to a convex set in $\mathbb{R}^{|\Omega|}$. Although this is not yet a convex program because of the bilinear equality constraint, we can convert it into one by letting $t_a(\omega) = m_a(\omega) b_a$, and noticing that $b_a \in [0,1]$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ if and only if $t_a \in \mathsf{Conv}(\widehat{\mathcal{P}}_a)$ where[9] $\widehat{\mathcal{P}}_a = \mathcal{P}_a \cup \{\mathbf{0}\}$. Using these expressions, we obtain the following main theorem:

Theorem 4.1. *The sender's persuasion problem* (4.3) *can be optimized by solving the following convex optimization problem:*

$$\begin{aligned}
&\max_{\{t_a : a \in A\}} \sum_{\omega \in \Omega} \sum_{a \in A} t_a(\omega) v(\omega, a) \\
&\textit{subject to,} \quad \sum_{a \in A} t_a(\omega) = \mu^*(\omega), \quad \textit{for each } \omega \in \Omega. \\
&\qquad t_a \in \mathsf{Conv}(\widehat{\mathcal{P}}_a), \quad \textit{for each } a \in A.
\end{aligned} \tag{4.9}$$

Proof. Any feasible $t_a \in \mathsf{Conv}(\widehat{\mathcal{P}}_a)$ can be written as $t_a = b_a m_a + (1 - b_a)\mathbf{0}$, for some $b_a \in [0,1]$ and $m_a \in \mathsf{Conv}(\mathcal{P}_a)$. Then, from Lemma 4.2, we obtain a corresponding Bayes-plausible η and an optimal strategy for the receiver $a(s)$, for which the sender's utility equals the objective of (4.9). □

Recall that the variable m_a denotes the *mean posterior* of the receiver, conditional on her choosing action a. Furthermore, b_a denotes the probability that the receiver chooses action a. Together this implies that $t_a(\omega) = b_a m_a(\omega)$ denotes the joint probability that the receiver

[9]Here, $\mathbf{0} \in \mathbb{R}^{|\Omega|}$ is the vector of all zeros.

takes action a and the realized state is $\bar{\omega} = \omega$. Thus, the reformulated problem (4.9) directly optimizes the joint probability distribution of the state and the receiver's actions.

Next, we describe how to get an optimal signaling scheme π from the optimal solution t_a to the problem (4.9). As in the proof of Theorem 4.1, for each $a \in A$, let $m_a \in \mathsf{Conv}(\mathcal{P}_a)$ be such that $t_a = b_a m_a$ for some $b_a \in [0,1]$. (Note that for $t_a \neq \mathbf{0}$, the corresponding m_a is uniquely defined.) As in the proof of Lemma 4.2, there exists $\{\mu_i^a : i = 1, \cdots, j_a\}$ and $\{\lambda_i^a : i = 1, \cdots, j_a\}$ for some $j_a \geq 1$, such that $\mu_i^a \in \mathcal{P}_a$, $\lambda_i^a \geq 0$, $\sum_{a \in A} \lambda_i^a = 1$, and $m_a = \sum_{i=1}^{j_a} \lambda_i^a \mu_i^a$. An optimal signaling scheme π is then given by the (discrete) distribution over $\Delta(\Omega \times \Delta(\Omega))$ that chooses $(\bar{\omega}, \bar{s}) = (\omega, \mu_i^a)$ with probability $b_a \lambda_i^a \mu_i^a(\omega)$. Note that conditional on $\bar{\omega} = \omega$, the optimal signaling scheme π makes the belief recommendation μ_i^a to the receiver with probability

$$\pi(\bar{s} = \mu_i^a | \bar{\omega} = \omega) = \frac{b_a \lambda_i^a \mu_i^a(\omega)}{\sum_{a' \in A} b_{a'} \lambda_i^{a'} \mu_i^{a'}(\omega)}.$$

We note that that in general, the representation of m_a as a convex combination of $\mu_i^a \in \mathcal{P}_a$ need not be unique. Since the preceding construction works for any convex decomposition of m_a, we conclude that there may exist multiple optimal signaling schemes for the receiver.

The preceding characterization also allows us to bound the size of the set of signals the sender needs to use to optimally persuade the receiver:

Theorem 4.2. *There exists an optimal signaling scheme $\pi \in \Delta(\Omega \times S)$, where the set S satisfies $|S| \leq |A| \cdot |\Omega|$. Specifically, for any $a \in A$, the signaling scheme sends at most $|\Omega|$ signals for which the receiver's optimal action is a.*

Proof. Observe that the set $\mathsf{Conv}(\mathcal{P}_a) \subseteq \Delta(\Omega)$ lies in an affine space of dimension $\mathbb{R}^{|\Omega|-1}$. This follows from the fact that $\Delta(\Omega) \subseteq \{x \in \mathbb{R}^{|\Omega|} : \sum_{\omega \in \Omega} x(\omega) = 1\}$. Since the optimal $m_a \in \mathsf{Conv}(\mathcal{P}_a)$, using Caratheodory's theorem (Bárány and Onn 1995), it follows that m_a can be written as a convex combination of at most $|\Omega|$ points in $\mathcal{P}_a$. That is, there exist

$\{\mu_i^a : i = 1, \cdots, j_a\}$ and $\{\lambda_i^a : i = 1, \cdots, j_a\}$ for some $j_a \leq |\Omega|$, such that $\mu_i^a \in \mathcal{P}_a$, $\lambda_i^a \geq 0$, $\sum_{a \in A} \lambda_i^a = 1$, and $m_a = \sum_{i=1}^{j_a} \lambda_i^a \mu_i^a$.

As detailed in the paragraph preceding the theorem statement, one can then construct an optimal signaling scheme using such a convex combination that sends $j_a \leq |\Omega|$ signals for which the receiver's optimal action is a, for each $a \in A$. Hence, the total number of signals is at most $|\Omega| \cdot |A|$. $\square$

Remark 4.1. *The preceding result can be strengthened as follows: For any set $A \in \mathbb{R}^{|\Omega|}$, let $\mathsf{Cara}(A)$ denote the minimum value of j such that any point $x \in \mathsf{Conv}(A)$ can be written as a convex combination of at most j points in A. Using the same arguments as in the proof of Theorem 4.2, we can show the existence of an optimal signaling scheme with at most $\sum_{a \in A} \mathsf{Cara}(\mathcal{P}_a)$ signals. Theorem 4.2 follows from Caratheodory's theorem which states $\mathsf{Cara}(A) \leq \dim(A)$ for any set A, where $\dim(A)$ is dimension of the smallest affine space containing A. Finally, note that $\mathsf{Cara}(A) = 1$ for any convex set A. Thus, in the case of expected utility maximizing agents, where all sets $\mathcal{P}_a$ are convex, we conclude that at most $|A|$ signals suffice to obtain an optimal signaling scheme.*

Finally, we briefly remark on the complexity of finding the optimal solution to the problem (4.9). Since the optimization problem is convex in the variables $\{t_a\}_{a \in A}$, this complexity rests on whether there exists an efficient characterization of the set $\mathsf{Conv}(\widehat{\mathcal{P}}_a)$ for each $a \in A$. Observe that this set is fully determined by the model primitives: in particular, the utility functions $\rho(\cdot, a)$ for each $a \in A$. Thus, whether an efficient characterization of the set $\mathsf{Conv}(\widehat{\mathcal{P}}_a)$ exists depends solely on the properties of the receiver's utility functions $\rho(\cdot, a)$ for each a. In the next section, we show that under some natural convexity properties on the utility function, one can replace the sets $\mathsf{Conv}(\widehat{\mathcal{P}}_a)$ by the convex hull of a finite number of pre-specified points. Using this, the sender's persuasion problem can be reduced to a linear program.

4.4 Binary persuasion

We now focus on a specific setting that is of practical importance. In this setting, the receiver's action space is binary, $A = \{0, 1\}$ and the sender's utility is always weakly higher when the receiver takes action 1: $v(\omega, 1) \geq v(\omega, 0)$ for all $\omega \in \Omega$. This model matches settings where, independent of the state, the sender seeks to persuade the receiver to engage with social media platforms (Candogan and Drakopoulos 2017), to join a queue (Lingenbrink and Iyer 2018a), or purchase a product (Lingenbrink and Iyer 2018b, Drakopoulos et al. 2018).

In the following, we define the receiver's *differential utility* $\bar{\rho}(\cdot)$ function as the difference in the utility between choosing action 1 and action 0:

$$\bar{\rho}(\mu) = \rho(\mu, 1) - \rho(\mu, 0).$$

Note that action $a = 1$ is optimal for the receiver at belief μ if and only if $\bar{\rho}(\mu) \geq 0$.

4.4.1 Geometry of the convex program

We see in the previous section that we can find the optimal signaling scheme as the solution to a convex program. However, the convex program had variables in $\mathsf{Conv}(\mathcal{P}_a \cup \{0\})$, for each action a. In this section, we show how this space can be reduced if we make a further assumption about the receiver utility.

Assumption 4.2. *The differential utility function, $\bar{\rho}(\mu)$, is convex.*

When $\rho(\mu, 1)$ is convex and $\rho(\mu, 0)$ is concave, this assumption holds. Of special note is letting $\rho(\mu, 0) = 0$, which is a common way of handling a receiver taking an outside option instead of making the sender-preferred action. Then, Assumption 4.2 requires that $\rho(\mu, 1)$ is convex. Intuitively, convexity of $\bar{\rho}$ means that the receiver dislikes uncertainty: the utility

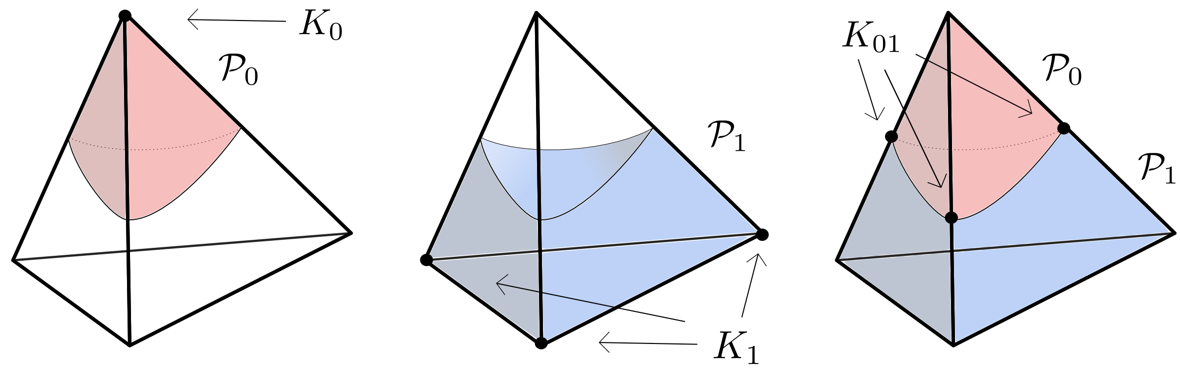


Figure 4.2: Geometry of K_0, K_1 and K_{01}.

of an uncertain outcome from a mixed distribution $\gamma\mu + (1-\gamma)\mu'$ is weakly lower than the weighted average of the utilities from the ingredients μ and μ'.

This assumption implies something significant about the geometry of the convex program (4.9): since $\mathcal{P}_0 = \{\mu : \bar{\rho}(\mu) \leq 0\}$, Assumption 4.2 implies that the set $\mathcal{P}_0$ is convex.

Using this assumption, we now show that (4.9) can be expressed as a linear program. In the following, to improve readability, we slightly abuse notation and let $\omega \in \Omega$ also denote the belief in $\Delta(\Omega)$ that assigns all weight to ω. Then, for each action $a \in A$, we let $K_a = \mathcal{P}_a \cap \Omega$ denote the states where the receiver prefers action a under full-information.[10] K_1 are the set of states where the receiver would take action 1 under full-information, and K_0 are the states where they would take action 0.

Next, we for $\omega_0 \in K_0$ and $\omega_1 \in K_1$, we define

$$\gamma(\omega_0, \omega_1) = \max_{\gamma \in [0,1]} \{\gamma : \bar{\rho}(\gamma\omega_0 + (1-\gamma)\omega_1) \geq 0\}.$$

Given an element $\omega_0 \in K_0$ and an element $\omega_1 \in K_1$, this quantity gives the greatest value amount of weight a belief could put on ω_0 (with the rest on ω_1) so that a receiver would take action 1. Since $\bar{\rho}(0) \geq 0$ and $\bar{\rho}(1) \leq 0$ and $\bar{\rho}$ is continuous, this function is well-defined. For

[10]If $\bar{\rho}(\omega) = 0$ for some $\omega \in \Omega$, then $\omega \in K_0 \cap K_1$. Hence, $K_0 \cap K_1$ may be non-empty.

any $\omega_0 \in K_0 \cap K_1^c$ and $\omega_1 \in K_1$, we define

$$\chi(\omega_0, \omega_1) = \gamma(\omega_0, \omega_1)\omega_0 + (1 - \gamma(\omega_0, \omega_1))\omega_1,$$

denote the belief over ω_0 and ω_1 for which the receiver is indifferent. Let K_{01} to be the set of these beliefs:

$$K_{01} = \{\chi(\omega_0, \omega_1) \in \Delta(\Omega) \text{ for some } \omega_0 \in K_0 \cap K_1^c, \omega_1 \in K_1\}. \tag{4.10}$$

With these definitions, we can state the main theorem of this section.

Theorem 4.3. *When $\bar{\rho}$ satisfies Assumption 4.2. the sender's persuasion problem can be optimized by solving the following linear program:*

$$\begin{aligned} &\max_{t_0, t_1} \sum_{\omega \in \Omega} (v(\omega, 1) - v(\omega, 0)) t_1(\omega) \\ \textit{subject to,} \quad & t_0 \in \mathsf{Conv}(K_0 \cup K_{01} \cup \{\mathbf{0}\}), \\ & t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\}), \\ & t_0(\omega) + t_1(\omega) = \mu^*(\omega) \quad \textit{for each } \omega \in \Omega. \end{aligned} \tag{4.11}$$

To prove this reduction to a linear program, we require two lemmas that describe the geometry of $\Delta(\Omega)$. The first lemma shows that the set of beliefs, $\Delta(\Omega)$, can be viewed as the union of two regions, each of which is the convex hull of a finite set of points. Its proof is given in Appendix C.1.1 and, unlike the theorem or the following lemma, does not require Assumption 4.2.

Lemma 4.3. $\Delta(\Omega) = \mathsf{Conv}(K_0 \cup K_{01}) \cup \mathsf{Conv}(K_1 \cup K_{01})$.

Figure 4.3 provides some geometric intuition for this lemma. Clearly, $\mathsf{Conv}(K_0 \cup K_{01}) \cup \mathsf{Conv}(K_1 \cup K_{01})$ is equal to $\Delta(\Omega)$ in this setting. However, it is generally not the case that the intersection $\mathsf{Conv}(K_0 \cup K_{01}) \cap \mathsf{Conv}(K_1 \cup K_{01}) = \mathsf{Conv}(K_{01})$ is a hyperplane. Figure C.1 in Appendix C.1.2 gives an example where $|K_0| > 1$ and $\dim(\mathsf{Conv}(K_{01})) = \dim(\Delta(\Omega))$.

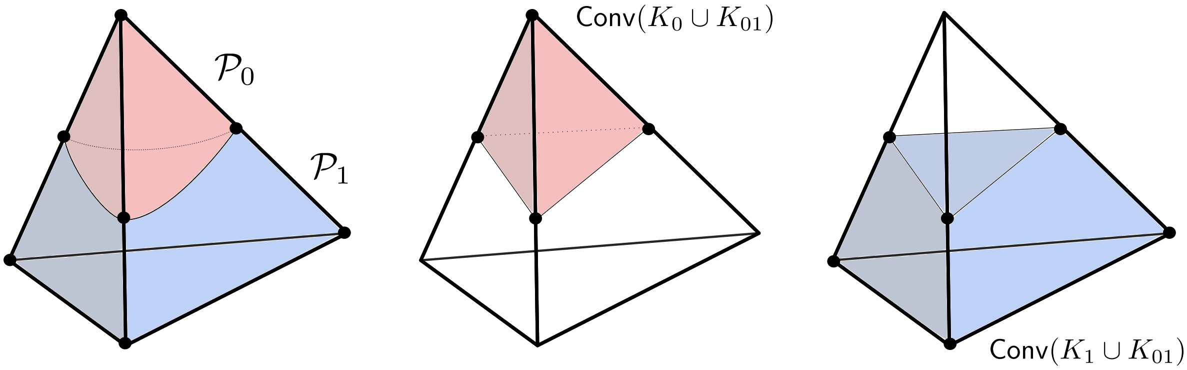


Figure 4.3: Geometry of $\Delta(\Omega)$.

We additionally make use of the following lemma, which says $\mathsf{Conv}(\mathcal{P}_1)$ is contained in only one of these regions: $\mathsf{Conv}(K_1 \cup K_{01})$. Since that region is itself a subset of $\mathsf{Conv}(\mathcal{P}_1)$, we have found a new expression for $\mathsf{Conv}(\mathcal{P}_1)$.

Lemma 4.4. *When $\bar{\rho}$ satisfies Assumption 4.2,* $\mathsf{Conv}(\mathcal{P}_1) = \mathsf{Conv}(K_1 \cup K_{01})$.

Proof. By Lemma 4.3, $\mathsf{Conv}(\mathcal{P}_1) \subset \mathsf{Conv}(K_0 \cup K_{01}) \cup \mathsf{Conv}(K_1 \cup K_{01})$. Suppose $\mu \in \mathcal{P}_1$ and $\mu \in \mathsf{Conv}(K_0 \cup K_{01})$. We can write

$$\mu = \sum_{\omega \in K_0} \alpha_\omega \omega + \sum_{\phi \in K_{01}} \alpha_\phi \phi.$$

By the convexity of $\bar{\rho}$,

$$\bar{\rho}(\mu) \leq \sum_{\omega \in K_0} \alpha_\omega \bar{\rho}(\omega) + \sum_{\phi \in K_{01}} \alpha_\phi \bar{\rho}(\phi) = \sum_{\omega \in K_0} \alpha_\omega \bar{\rho}(\omega) + 0.$$

For any $\omega \in K_0$, unless $\omega \in K_0 \cap K_1$, $\bar{\rho}(\omega) < 0$. Since $\mu \in \mathcal{P}_1$, we have $\bar{\rho}(\mu) \geq 0$, and hence $\alpha_\omega = 0$ unless $\omega \in K_1$. Thus, $\mu \in \mathcal{P}_1$ implies $\mu \in \mathsf{Conv}(K_1 \cup K_{01})$, and hence $\mathsf{Conv}(\mathcal{P}_1) = \mathsf{Conv}(K_1 \cup K_{01})$. ◻

Finally, with these lemmas, we can prove Theorem 4.3.

Proof of Theorem 4.3. Since $t_0(\omega) + t_1(\omega) = \mu^*(\omega)$ for each $\omega \in \Omega$, it is straightforward to verify that the objectives of (4.9) and (4.11) differ only by the constant term $\sum_{\omega \in \Omega} \mu^*(\omega) v(\omega, 0)$.

Thus, to prove the result, it suffices to show that optimal solution of each program is a feasible for the other.

First, consider any optimal solution to (4.11). Since $\mathsf{Conv}(K_0 \cup K_{01} \cup \{\mathbf{0}\}) \subset \mathsf{Conv}(\widehat{\mathcal{P}}_0)$ and $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\}) \subset \mathsf{Conv}(\widehat{\mathcal{P}}_1)$, it is a feasible solution to (4.9).

Now, consider any optimal solution to (4.9), $\{t_0, t_1\}$, where $t_0 = b_0 m_0 + (1 - b_0)\mathbf{0}$ and $t_1 = b_1 m_1 + (1 - b_1)\mathbf{0}$, where $m_0 \in \mathsf{Conv}(\mathcal{P}_0)$, $m_1 \in \mathsf{Conv}(\mathcal{P}_1)$ and $b_0, b_1 \in [0, 1]$. By Lemma 4.4, $m_1 \in \mathsf{Conv}(K_1 \cup K_{01})$ and hence $t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$. By Lemma 4.3, $m_0 \in \mathsf{Conv}(K_0 \cup K_{01}) \cup \mathsf{Conv}(K_1 \cup K_{01})$. Suppose $m_0 \in \mathsf{Conv}(K_1 \cup K_{01})$ and $b_0 \neq 0$. Thus, $t_0 \neq \mathbf{0}$ and the solution has objective value less than 1. Further, $\mu^* = b_0 m_0 + b_1 m_1 \in \mathsf{Conv}(K_1 \cup K_{01})$ and the solution $\hat{t}$ to (4.9), where $\hat{t}_0 = \mathbf{0}$ and $\hat{t}_1 = t_0 + t_1$ achieves revenue 1. This contradicts the optimality of t. Thus, if $m_0 \in \mathsf{Conv}(K_1 \cup K_{01})$, then $b_0 = 0$ and $t_0 = 0$. Hence, we conclude either $b_0 = 0$ or $\mu_0 \in \mathsf{Conv}(K_0 \cup K_{01})$, and thus $t_0 \in \mathsf{Conv}(K_0 \cup K_{01} \cup \{\mathbf{0}\})$, and the solution is a feasible solution of (4.11). □

4.4.2 Structural Characterizations

The formulation of linear program (4.3) allows the optimal signaling scheme to be found much easier. Further specification of a Bayesian persuasion problem may allow that linear program to be further reduced. Additionally, Theorem 4.3 provides several implications about the structure of the optimal signaling scheme.

Theorem 4.3 suggests there is a canonical set of signals that any optimal signaling scheme will use: $K_1 \cup K_0 \cup K_{01}$ (when $\mu^* \notin P_1$, $K_0 \cup K_{01}$). To see this, recall that (4.3) restricted t_0 and t_1 to $\mathsf{Conv}(K_0 \cup K_{01} \cup \{\mathbf{0}\})$ and $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, respectively. Then, the mean posteriors m_0 and m_1 are in $\mathsf{Conv}(K_0 \cup K_{01}\})$ and $\mathsf{Conv}(K_1 \cup K_{01})$, respectively. This implies that m_0 can be fully expressed as the mean of signals in $K_1 \cup K_{01}$, and similarly m_1 can be

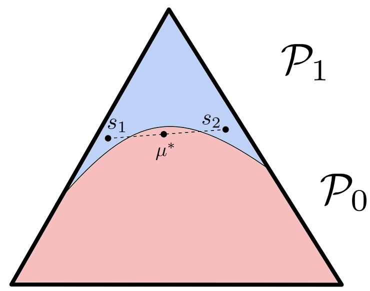


Figure 4.4: Receiver is fully persuaded.

expressed as the mean of signals in $K_0 \cup K_{01}$. There then exists a sender utility-optimal signaling scheme where a receiver is always informed either the exact state ω or a set of two states $\{\omega_0, \omega_1\}$ the true state is in. Of course, by the convexity of $\mathcal{P}_0$, there also exists a signaling mechanism where all posteriors that lead to a receiver are replaced with a "take action 0" signal. We can further characterize the signals that lead to the receiver taking action $a = 0$ with the following proposition.

Proposition 4.1 (worst action for sender implies receiver is certain)**.** *If, in an optimal signaling scheme, a receiver's belief μ leads to her taking action $a = 0$, then μ assigns zero weight to all elements of K_1.*

This proposition is a corollary of Proposition 4 in Kamenica and Gentzkow (2011) and implies that the signals that lead to a receiver choosing action 0 must assign positive weights only to states in K_0. Further, when $\mu^* \notin \mathsf{Conv}(P_1)$ and when $\bar{\rho}$ is strictly convex, one can further show that in any optimal signaling scheme, the signals that lead to a receiver choosing action 1 must exactly correspond to beliefs in K_{01}, and not in K_1.

Counterintuitively, the convexity of $\mathcal{P}_0$ and non-convexity of $\mathcal{P}_1$ allows the sender to sometimes fully persuade receivers when the prior is in $\mathcal{P}_0$. For example, in Figure 4.4, we have $\mu^* \in \mathsf{Conv}(\mathcal{P}_0)$. However, we can find beliefs $s_1, s_2 \in \mathcal{P}_1$ such that $\gamma s_1 + (1-\gamma)s_2 = \mu^*$.

The signaling scheme that induces posterior s_1 with probability γ and s_2 with probability $1-\gamma$ is Bayes-plausible, and thus is a valid signaling scheme. Under this signaling scheme, at both s_1 and s_2, the receiver strictly prefers $a = 1$. Hence, the sender can fully persuade the receiver with the receiver strictly preferring $a = 1$ under both posteriors, even though she would have chosen $a = 0$ under her prior μ^*. Furthermore, this differs from Proposition 5 of Kamenica and Gentzkow (2011) which states that with an expected utility maximizing receiver, whenever the receiver takes $a = 1$, then the receiver is indifferent between at least two actions.

4.5 Signaling in Queues

To illustrate our methodology, we now apply our methods to setting of an unobservable $M/M/1$ queue where arriving customers must choose between joining or balking after being sent a signal by the principal. Customers frequently treat uncertain wait times as worse than longer, certain wait times, as empirically shown by Maister et al. (1984). Hence the question of how to best persuade customers to join a queue is an good application of our results.

4.5.1 Queueing Model

Like in Lingenbrink and Iyer (2018a), a service provider sees potential customers who arrive according to a Poisson process of rate λ. The service provider is limited on the rate they can provide service (they have a single server that takes time exponentially distributed at rate 1), so customers who arrive when the server is busy may wait in a first-in first-out queue to attain service. Upon arrival, each customer decides whether to wait in the queue or leave the system without obtaining service. We consider the case where customers cannot observe the queue length before making their decision. Instead, the service provider can observe the length and communicate this information to arriving customers. For this setting, the set of states are the set of possible queue lengths, and the customer's payoff-relevant random variable is their wait time. The service provider aims to maximize throughput through the system.

There are two key differences between the settings we have considered and this $M/M/1$ queue setting: an endogenous prior and an infinite set of states Ω. The endogenous prior changes the constraint $t_0(\omega) + t_1(\omega) = \mu^*(\omega)$ for all ω in the convex optimization program (4.9) to $\lambda t_1(\omega) = t_0(\omega + 1) + t_1(\omega + 1)$ for all ω.

Unlike the model we have used, the set of non-negative numbers is not finite. The proof provided of Lemma 4.3 depended on the finiteness of Ω. However, with the additional structure that the queue setting provides (notably that K_1 is finite), we can prove the result regardless. In fact, we can prove an analogous theorem to Theorem 4.3. The full argument is given in Appendix C.1.3.

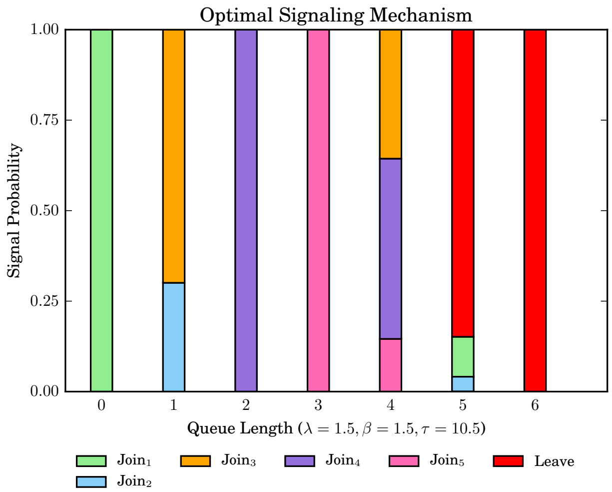
We can show a very similar result to Lingenbrink and Iyer (2018a). The optimal signaling scheme has a very specific form.

Theorem 4.4. *If Assumption 4.2 is satisfied and K_1 is finite, the optimal signaling scheme is a threshold mechanism: there exists a $x = m + \epsilon$ such that*

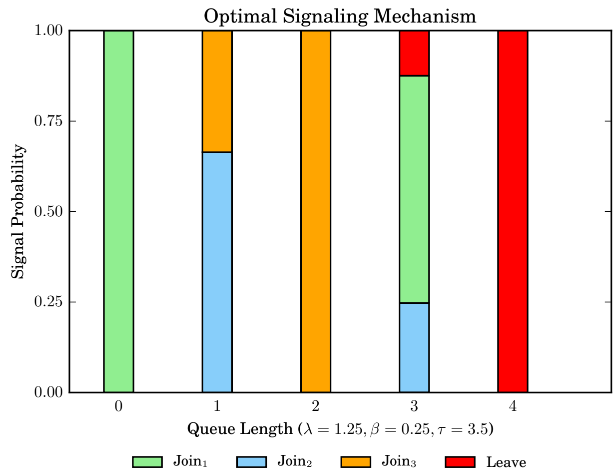
$$t_1(n) = \begin{cases} \lambda^n t & \text{if } 0 \leq n \leq m-1, \\ \epsilon \lambda^m t & \text{if } n = m, \\ 0 & \text{if } n \geq m+1, \end{cases}$$

$t_0(0) = 0$ *and* $t_0(n) = \lambda t_1(n-1) - t_1(n)$ *for* $n \geq 1$. *In particular, the customer joins with probability* 1 *if* $n \leq m$, ϵ *if* $n = m + 1$, *and* 0 *if* $n \geq m + 2$.

The proof of this theorem is given in Appendix C.1.4. Below, we give numerical results for a specific receiver utility function.



(a) $(\lambda, \beta, \tau) = (1.5, 0.5, 4.5)$



(b) $(\lambda, \beta, \tau) = (1.25, 0.25, 3.5)$

Figure 4.5: Optimal Signaling Mechanisms.

4.5.2 Numerical Examples

In this section, we give numerical examples of the optimal signaling schemes for $M/M/1$ queues and their associated optimal throughputs. We model customers as having utility given by the mean-stdev risk measure. That is, for some $\beta > 0$, let the differential utility function be

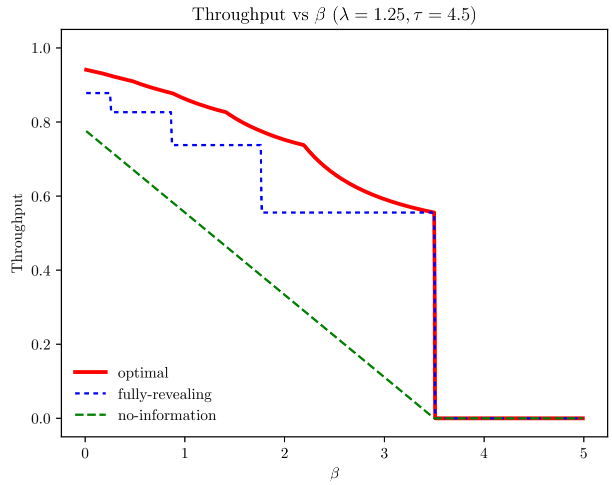
$$\bar{\rho}(\mu) = \tau - \left(\mathbf{E}_\mu[X] + \beta\sqrt{\text{Var}_\mu[X]}\right) \tag{4.12}$$

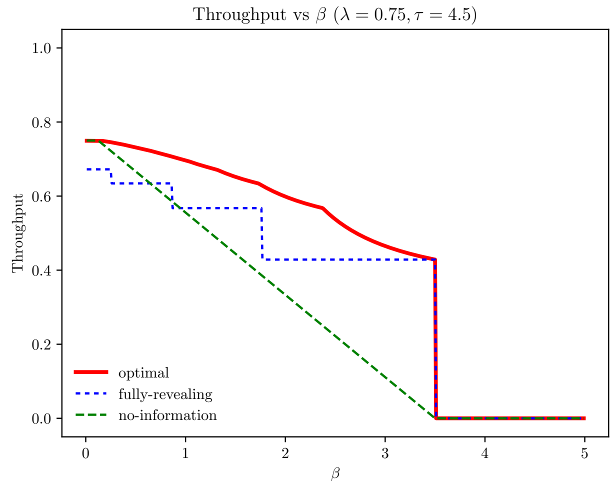
where X is a customer's waiting time and μ is their posterior over ω. We first show that this differential utility function satisfies Assumption 4.2.

Lemma 4.5. *$\bar{\rho}(\mu)$ is convex in μ.*

The proof is given in Appendix C.1.5. With this lemma, Theorem 4.4 applies.

Figure 4.5 shows the optimal signaling mechanisms for $(\lambda, \beta, \tau) = (1.5, 0.5, 4.5)$ and $(1.25, 0.25, 3.5)$ respectively. Each vertical bar shows the probability of sending particular signals conditional on queue length. For example, in Figure 4.5a, there are 4 join signals, $\text{Join}_1, \text{Join}_2, \text{Join}_3, \text{Join}_4$ and one Leave signal. When the queue length is 4, the queue sends





(a) Throughput versus β for $\lambda = 1.25$ and $\tau = 4.5$.(b) Throughput versus β for $\lambda = 0.75$ and $\tau = 4.5$.

Figure 4.6: Comparison of the optimal, the fully-revealing, and the no-information mechanisms.

Join_1 with probability 0.233, Join_2 with probability 0.185, and Leave with probability 0.582. Other signal probabilities conditional on queue length can be read from the plot analogously.

We can see from Figure 4.5 that the optimal mechanism is a threshold mechanism: in both examples, when the queue length is at most 3 the customer always joins, when the queue length is 4 the customer joins with probability in $[0, 1]$ and when the queue length is higher than 4 the customer never joins. We formalize and prove this observation in Theorem 4.4.

Figure 4.5 also suggests that under the optimal signaling mechanism, the join signals have a "sandwich" structure. Namely, the join signals can be ordered $\text{Join}_1, \ldots, \text{Join}_J$ such that for each $j \in [J]$, the posterior under Join_j puts weight on at most two queue lengths a_j, b_j, $a_j \leq b_j$, such that $a_1 \leq \cdots \leq a_J$ and $b_1 \geq \cdots \geq b_J$. Moreover, the signals are ordered by riskiness: $\mathbf{E}[X|\text{Join}_1] \leq \cdots \leq \mathbf{E}[X|\text{Join}_J]$ and $\text{Var}(X|\text{Join}_1) \geq \cdots \geq \text{Var}(X|\text{Join}_J)$ such that $\mathbf{E}[X|\text{Join}_j] + \beta\sqrt{\text{Var}(X|\text{Join}_j)} = \tau$ for all $j \in [J]$. Join_1 is the most risky signal with lowest expected waiting time and highest variability, and Join_J is the least risky signal with highest expected waiting time and lowest variability, but all join signals give the customer the same utility equal to her outside option.

For completeness, we compare the queue throughput under the optimal signaling mechanism with throughputs under two benchmarks: fully-revealing mechanism, where the queue operator reveals the queue length exactly to each arriving customer, and no-information, where the queue operator reveals no information to each arriving customer. The throughputs under fully-revealing and no-information mechanisms can be computed analytically; the details are given in Proposition C.1 in Appendix C.1.6.

We can see in Figure 4.6 that under all mechanisms, as β increases, the throughput decreases: when customers are more risk-conscious, they are less likely to be satisfied with the queue and join relative to their outside option. Up to a point, persuasion can increase throughput above what can be achieved under fully-revealing or no-information mechanisms, but when the risk-consciousness parameter β is high enough, no customer wants to join the queue under any mechanism because the uncertainty from its own service time alone is unacceptable, and persuasion no longer increases the sender's utility.

4.6 Discussion

We consider a general Bayesian persuasion with risk-conscious agents. We find that, unlike the case with an expected utility maximizing receiver, the sender often requires multiple signals to persuade the receiver to take an action. We show that the optimal signaling scheme can be found as the solution of a convex program and give a bound on the number of signals necessary. In the setting where the receiver can choose only two actions where one of which is always preferred by the sender, we find that this convex program can be expressed as a linear program under a mild convexity assumption on the receiver utility. Finally, we consider signaling in queues and obtain a characterization of the structure of the optimal signaling scheme.

In the section on binary persuasion, our results depended on the assumption that $\bar{\rho}(\mu) = \rho(\mu, 1) - \rho(\mu, 0)$ is convex. Finding the optimal signaling scheme under other forms of $\bar{\rho}$ remains an interesting problem. A natural question would be when the function is concave, modeling instances where the sender is persuading the receiver to take a risky action, and the receiver is risk-seeking. Note that, a concave $\bar{\rho}$ implies a convex $\mathcal{P}_1$. This implies the optimal signaling scheme need only send one signal to encourage customers to take action $a = 1$. However, it is difficult to reduce this problem to a linear program like we did with convex $\bar{\rho}$; $\mathcal{P}_1$ is convex but is not in general the convex hull of a finite (or countable) set of points.

In this chapter, we consider persuading a "receiver" who is a single agent and is not an expected utility maximizer. We might consider another formulation of this problem: *publicly persuading a group of agents.* Here, the "receiver" is a group of agents, each with separate (possibly expected) utilities. Any $a \in A$ denotes an equilibrium among the group of agents: $a = (a_1, a_2, \ldots)$ where a_i is the action taken by agent i. The set $\mathcal{P}_a$ then denotes when equilibrium a is preferred by the group of agents. Then, we can view this as the sender publicly signaling to the group of agents to convince them to more frequently play sender-preferred equilibriums.

Additionally, while we do not treat the receiver as necessarily being an expected utility maximizer, we do treat the sender as one. An interesting direction for further research would be to see what occurs when this modeling assumption is relaxed.

CHAPTER 5
CONCLUSION

> Ninety percent of selling is conviction, and 10 percent is persuasion.
>
> — Shiv Khera

This thesis applies information design and Bayesian persuasion to settings common to operations research: service systems and online markets. For readability, several modeling assumptions are made that are not necessary for the results. In Chapters 2 and 3, we use revenue as the firm's objective. However, our methods apply to many other forms of firm utility. Similarly, while Chapter 3 considers a specific model of retail, the results apply to many other information design problems that satisfy the mentioned assumptions about customer utility.

The results presented in these chapters depend on the assumption that the firm could commit to a manner of signaling. Realistically, there exist firms that operate in bad faith and may signal differently than promised. One might consider a repeated game where a firm has a fixed probability of being "credible," and having the ability to commit. During each repetition of the game, the firm observes the hidden state of the game, a new customer arrives, and the firm signals to a new customer an action recommendation based on the state. A credible firm can commit to a manner of signaling and follows it, but a non-credible firm cannot and always sends a signal that maximizes their utility. After each occurrence of the game, the recommendation of the firm and the realization of the state is revealed, and future customers will have a new belief on the credibility of the firm. A fruitful direction for future research would be finding how a credible firm can simultaneously establish their credibility and generate the maximum revenue.

Bibliography

Ahmadi-Javid, A. (2012). Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123.

Alizamir, S., de Véricourt, F., and Wang, S. (2018). Warning against recurring risks: An information design approach.

Allon, G. and Bassamboo, A. (2011). Buying from the babbling retailer? the impact of availability information on customer behavior. *Management Science*, 57(4):713–726.

Allon, G., Bassamboo, A., and Gurvich, I. (2011). We will be right with you: Managing customer expectations with vague promises and cheap talk. *Operations Research*, 59(6):1382–1394.

Anunrojwong, J., Iyer, K., and Lingenbrink, D. (2019). Persuading risk-conscious agents: A geometric approach. *SSRN Electronic Journal*.

Armony, M. and Maglaras, C. (2004a). Contact centers with a call-back option and real-time delay information. *Operations Research*, 52(4):527–545.

Armony, M. and Maglaras, C. (2004b). On customer contact centers with a call-back option: Customer decisions, routing rules, and system design. *Operations Research*, 52(2):271–292.

Ata, B., Glynn, P. W., and Peng, X. (2017). An equilibrium analysis of a discrete-time markovian queue with endogenous abandonments. *Queueing Systems*, 86(1):141–212.

Ata, B. and Peng, X. (2017). An equilibrium analysis of a multiclass queue with endogenous abandonments in heavy traffic. *Operations Research*, (forthcoming).

Aviv, Y., Wei, M. M., and Zhang, F. (2018). Responsive pricing of fashion products: The effects of demand learning and strategic consumer behavior.

Aydinliyim, T., Pangburn, M. S., and Rabinovich, E. (2017). Inventory disclosure in online retailing. *European Journal of Operational Research*, 261(1):195 – 204.

Bárány, I. and Onn, S. (1995). Carathéodory's theorem, colourful and applicable. *Intuitive Geometry*, 6:11–21.

Bergemann, D., Brooks, B., and Morris, S. (2015). The limits of price discrimination. *American Economic Review*, 105(3):921–57.

Bergemann, D. and Morris, S. (2016a). Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics*, 11(2):487–522.

Bergemann, D. and Morris, S. (2016b). Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium. *American Economic Review*, 106(5):586–591.

Bergemann, D. and Morris, S. (2018). Information design: A unified perspective. *Journal of Economic Literature*, Forthcoming.

Borgs, C., Chayes, J. T., Doroudi, S., Harchol-Balter, M., and Xu, K. (2014). The optimal admission threshold in observable queues with state dependent pricing. *Probability in the Engineering and Informational Sciences*, 28(1):101–119.

Burnetas, A. and Economou, A. (2007). Equilibrium customer strategies in a single server markovian queue with setup times. *Queueing Systems*, 56(3):213–228.

Candogan, O. and Drakopoulos, K. (2017). Optimal signaling of content accuracy: Engagement vs. misinformation. *Available at SSRN: https://ssrn.com/abstract=3051275*.

Carlo Acerbi, D. T. (2002). Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes*, 31(2):379–388.

Chen, H. and Frank, M. Z. (2001). State dependent pricing with a queue. *IIE Transactions*, 33(10):847–860.

Cominetti, R. and Torrico, A. (2016). Additive consistency of risk measures and its application to risk-averse routing in networks. *Mathematics of Operations Research*, 41(4):1510–1521.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6):1431–1451.

Cui, R. and Shin, H. (2018). Sharing aggregate inventory information with customers: Strategic cross-selling and shortage reduction. *Management Science*, 64(1):381–400.

Cui, R., Zhang, D., and Bassamboo, A. (2018). Learning from inventory availability information: Evidence from field experiments on amazon. *Management Science*.

Cui, S. and Veeraraghavan, S. (2016). Blind queues: The impact of consumer beliefs on revenues and congestion. *Management Science*, 62(12):3656–3672.

Doval, L. and Ely, J. (2016). Sequential information design. *Working paper*.

Drakopoulos, K., Jain, S., and Randhawa, R. (2018). Persuading customers to buy early: The value of personalized information provisioning. *Available at SSRN: https://ssrn.com/abstract=3191629*.

Dughmi, S. (2017). Algorithmic information structure design: A survey. *SIGecom Exch.*, 15(2):2–24.

Dughmi, S. and Xu, H. (2017). Algorithmic persuasion with no externalities. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, EC '17, pages 351–368.

Economou, A. and Kanta, S. (2008a). Equilibrium balking strategies in the observable single-server queue with breakdowns and repairs. *Operations Research Letters*, 36(6):696–699.

Economou, A. and Kanta, S. (2008b). Optimal balking strategies and pricing for the single server markovian queue with compartmented waiting space. *Queueing Systems*, 59(3):237.

Edelson, N. M. and Hilderbrand, D. K. (1975). Congestion tolls for poisson queuing processes. *Econometrica*, 43(1):81–92.

Ely, J. C. (2017). Beeps. *American Economic Review*, 107(1):31–53.

Feld, S. L. (1991). Why your friends have more friends than you do. *American Journal of Sociology*, 96(6):1464–1477.

Föllmer, H. and Schied, A. (2011). *Stochastic finance: an introduction in discrete time*. Walter de Gruyter.

Fudenberg, D. and Tirole, J. (1991). *Game theory*. The MIT Press, Cambridge, Massachusetts.

Garnett, O., Mandelbaum, A., and Reiman, M. (2002). Designing a call center with impatient customers. *Manufacturing & Service Operations Management*, 4(3):208–227.

Genesove, David; Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics*, 116(4):1233–1260.

Gentzkow, M. and Kamenica, E. (2016). A rothschild-stiglitz approach to bayesian persuasion. *American Economic Review*, 106(5):597–601.

Guo, P. and Zipkin, P. (2007). Analysis and comparison of queues with different levels of delay information. *Management Science*, 53(6):962–970.

Hassin, R. (1986). Consumer information in markets with random product quality: The case of queues and balking. *Econometrica*, 54(5):1185–1195.

Hassin, R. (2016). *Rational queueing.* CRC Press, Taylor & Francis Group.

Hassin, R. and Haviv, M. (2012). *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems.* International Series in Operations Research & Management Science. Springer US.

Hassin, R. and Koshman, A. (2017). Profit maximization in the M/M/1 queue. *Operations Research Letters*, 45:436–441.

Hassin, R. and Roet-Green, R. (2017). The impact of inspection cost on equilibrium, revenue, and social welfare in a single-server queue. *Operations Research*, 65(3):804–820.

Jiang, B., Tian, L., Xu, Y., and Zhang, F. (2016). To share or not to share: Demand forecast sharing in a distribution channel. *Marketing Science*, 35(5):800–809.

Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk.* McGraw-Hill, New York.

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3(3):430–454.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kalwani, M. U., Yim, C. K., Rinne, H. J., and Sugita, Y. (1990). A price expectations model of customer brand choice. *Journal of Marketing Research*, 27(3):251–262.

Kamenica, E. (2018). Bayesian persuasion and information design. *Annual Review of Economics*, Forthcoming.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kolotilin, A., Mylovanov, T., Zapechelnyuk, A., and Li, M. (2016). Persuasion of a privately informed receiver. *UNSW Business School Research Paper*, pages No. 2016–21.

Korte, B. and Vygen, J. (2012). *Combinatorial Optimization: Theory and Algorithms.* Algorithms and Combinatorics. Springer Berlin Heidelberg.

Kremer, I., Mansour, Y., and Perry, M. (2014). Implementing the “wisdom of the crowd”. *Journal of Political Economy*, 122(5):988–1012.

Lingenbrink, D. and Iyer, K. (2018a). Optimal signaling mechanisms in unobservable queues. *Operations Research*, Forthcoming.

Lingenbrink, D. and Iyer, K. (2018b). Signaling in online retail: Efficacy of public signals. *SSRN Electronic Journal*.

Machina, M. J. (1995). Non-expected utility and the robustness of the classical insurance paradigm. *The Geneva Papers on Risk and Insurance Theory*, 20(1):9–50.

Maister, D. H. et al. (1984). *The psychology of waiting lines*. Harvard Business School, Boston.

Mansour, Y., Slivkins, A., Syrgkanis, V., and Wu, Z. S. (2016). Bayesian exploration: Incentivizing exploration in bayesian games. *CoRR*, abs/1602.07570.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

McAfee, R. and McMillan, J. (1987). Auctions with a stochastic number of bidders. *Journal of Economic Theory*, 43(1):1 – 19.

Morris, S. and Shin, H. S. (2003). *Global Games: Theory and Applications*, volume 1 of *Econometric Society Monographs*, page 56âĂŞ114. Cambridge University Press.

Naor, P. (1969). The regulation of queue size by levying tolls. *Econometrica*, 37(1):15–24.

Nikolova, E. (2010). Approximation algorithms for reliable stochastic combinatorial optimization. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 338–351. Springer, Berlin.

Nikolova, E. and Stier-Moses, N. E. (2014). A mean-risk model for the traffic assignment problem with stochastic travel times. *Operations Research*, 62(2):366–382.

Nikolova, E. and Stier-Moses, N. E. (2015). The burden of risk aversion in mean-risk selfish routing. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 489–506, Portland, Oregon. ACM.

Papanastasiou, Y., Bimpikis, K., and Savva, N. (2017). Crowdsourcing exploration. *Management Science*, (forthcoming).

Peinkofer, S. T., Esper, T. L., and Howlett, E. (2016). Hurry! sale ends soon: The impact of limited inventory availability disclosure on consumer responses to online stockouts. *Journal of Business Logistics*, 37(3):231–246.

Pender, J., Rand, R. H., and Wesson, E. (2017). Queues with choice via delay differential equations. *International Journal of Bifurcation and Chaos*, 27(04):1730016.

Pender, J., Rand, R. H., and Wesson, E. (2018). An analysis of queues with delayed information and time-varying arrival rates. *Nonlinear Dynamics*, 91(4):2411–2427.

Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, 36(1):11–46.

Rayo, L. and Segal, I. (2010). Optimal information disclosure. *Journal of Political Economy*, 118(5):949–987.

Romanyuk, G. and Smolin, A. (2018). Cream skimming and information design in matching markets. *American Economic Journal: Microeconomics*, Forthcoming.

Sheshinski, E. and Weiss, Y. (1977). Inflation and costs of price adjustment. *The Review of Economic Studies*, 44(2):287–303.

Simhon, E., Hayel, Y., Starobinski, D., and Zhu, Q. (2016). Optimal information disclosure policies in strategic queueing games. *Operations Research Letters*, 44(1):109–113.

Taneva, I. (2019). Information design. *American Economic Journal: Microeconomics*, Forthcoming.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26(1):71–92.

Wolff, R. W. (1982). Poisson arrivals see time averages. *Operations Research*, 30(2):223–231.

Yu, Q., Allon, G., and Bassamboo, A. (2017a). How do delay announcements shape customer behavior? an empirical study. *Management Science*, 63(1):1–20.

Yu, Q., Allon, G., Bassamboo, A., and Iravani, S. (2017b). Managing customer expectations and priorities in service systems. *Management Science*, (forthcoming).

APPENDIX A

APPENDIX TO CHAPTER 2

A.1 Proofs

In this section, we provide the proofs of the results in the main body of the chapter. We start with the proofs of the lemmas in Section 2.3.

Proof of Lemma 2.1. Given a signaling mechanism $\Sigma = (\mathcal{S}, \sigma)$ and customer equilibrium f, consider a new signaling mechanism $\Sigma_1 = (\mathcal{S}_1, \sigma_1)$, where $\mathcal{S}_1 = \mathcal{S} \times \{0, 1\}$, and $\sigma_1 : \mathbb{N}_0 \times \mathcal{S}_1 \to [0, 1]$ is given by

$$\sigma_1(n, s, 1) = \sigma(n, s) f(s)$$
$$\sigma_1(n, s, 0) = \sigma(n, s)(1 - f(s)).$$

Now consider the strategy f_1 under the signaling mechanism Σ_1, where $f_1(s, 1) = 1$ and $f_1(s, 0) = 0$. We begin by showing that the strategy f_1 constitutes a customer equilibrium under Σ_1. First, note that the steady state distribution of the queue under (Σ_1, f_1) is same as that under (Σ, f). This follows from the fact that the queue has the same transition probabilities at each state under the two settings. Denote this common steady state distribution by π_∞. From this, we obtain

$$\begin{aligned}
\mathbf{E}^{\Sigma_1, f_1}[h(X_\infty, p)|(s, 1)] &= \sum_{n=0}^{\infty} \pi_\infty(n|s, 1) h(n, p) \\
&= \sum_{n=0}^{\infty} \pi_\infty(n|s) h(n, p) \\
&= \mathbf{E}^{\Sigma, f}[h(X_\infty, p)|s].
\end{aligned}$$

Here, $\pi_\infty(n|s, 1)$ and $\pi_\infty(n|s)$ denote the conditional probability that there are n customers in the queue upon seeing a signal $(s, 1)$ and s respectively in the two signaling mechanisms.

The second equality follows from the fact that the choice of the second component in σ_1 is independent of the number of customers in the queue.

Note that under Σ_1, a customer sees the signal $(s,1)$ only if $f(s) > 0$, which implies, from the fact that f is a customer equilibrium under σ, that $\mathbf{E}^{\Sigma,f}[h(X_\infty,p)|s] \geq 0$. This implies that $\mathbf{E}^{\Sigma_1,f_1}[h(X_\infty,p)|(s,1)] \geq 0$, and indeed $f_1(s,1) = 1$ is an optimal action. Similarly, we obtain that if the customer observes the signal $(s,0)$ then $f_1(s,0) = 0$ is indeed an optimal action. Together, we obtain that the strategy f_1 is a customer equilibrium under Σ_1.

The proof then follows from the fact that f_1 is a pure strategy. $\square$

Proof of Lemma 2.2. From Lemma 2.1, without loss of generality, assume that the customer equilibrium f is pure. Let $\mathcal{S}^i = \{s \in \mathcal{S} : f(s) = i\}$ for $i = 0,1$. Define $\sigma_1 : \mathbb{N} \times \{0,1\} \to [0,1]$ as follows

$$\sigma_1(n,i) = \sum_{s \in \mathcal{S}^i} \sigma(n,s), \quad \text{for } i = 0,1.$$

Now consider the strategy f_1 under the signaling mechanism σ_1, where $f_1(i) = i$ for $i = 0,1$. By similar argument as in Lemma 2.1, it follows that the steady state distribution under (Σ, f) is same as that under (Σ_1, f_1). Denote this steady state distribution by π_∞. Thus, it follows that the two settings are equivalent, if we show that f_1 is indeed a customer equilibrium under Σ_1. This follows directly by observing that $\pi_\infty(n|i=1) = \pi_\infty(n|s \in \mathcal{S}^1)$, and hence

$$\begin{aligned}
\mathbf{E}^{\Sigma_1,f_1}[h(X_\infty,p)|i=1] &= \mathbf{E}^{\Sigma,f}[h(X_\infty,p)|s \in \mathcal{S}^1] \\
&= \sum_{n=0}^{\infty} \pi_\infty(n|s \in \mathcal{S}^1) h(n,p) \\
&= \sum_{s \in \mathcal{S}^1} \frac{\pi_\infty(s)}{\pi_\infty(\mathcal{S}^1)} \sum_{n=0}^{\infty} \pi_\infty(n|s) h(n,p) \\
&= \frac{1}{\pi_\infty(\mathcal{S}^1)} \sum_{s \in \mathcal{S}^1} \pi_\infty(s) \;\; \mathbf{E}^{\Sigma,f}[h(X_\infty,p)|s] \\
&\geq 0,
\end{aligned}$$

since $\mathbf{E}^{\Sigma,f}[h(X_\infty, p)|s] \geq 0$ for all $s \in \mathcal{S}^1$. Here, $\pi_\infty(s)$ denotes the probability in steady state of seeing signal s upon arrival, and $\pi_\infty(\mathcal{S}^1)$ denotes the probability of seeing a signal in $\mathcal{S}^1$. The third equation follows from the law of total probability. From this, we obtain that $f_1(1) = 1$ is an optimal action on observing a signal 1 under σ_1. Similarly, we obtain that $f(0) = 0$ is an optimal action on observing 0 under σ_1. Together, this implies that f_1 is a customer equilibrium under Σ_1 and the result follows. $\square$

Proof of Lemma 2.3. We begin by showing that for any signaling mechanism $\sigma : \mathbb{N}_0 \times \mathcal{S} \to [0, 1]$ feasible for (2.4), there exists a feasible solution $\pi = \{\pi_n : n \geq 0\}$ to the linear program (2.5) with the same objective value. Note that the steady state distribution π_∞^σ of the queue under σ in the obedient equilibrium satisfies the following detailed balance equation,

$$\pi_\infty^\sigma(n)\lambda\sigma(n, 1) = \pi_\infty^\sigma(n + 1),$$

implying that

$$\pi_\infty^\sigma(n) = \lambda^n \left(\prod_{j=0}^{n-1} \sigma(j, 1)\right) \pi_\infty^\sigma(0), \tag{A1}$$

with $\pi_\infty^\sigma(0)$ given by

$$\pi_\infty^\sigma(0) = \left(\sum_{n=0}^{\infty} \lambda^n \left(\prod_{j=0}^{n-1} \sigma(j, 1)\right)\right)^{-1}. \tag{A2}$$

Define π as $\pi_n = \pi_\infty^\sigma(n)$ for all $n \geq 0$. Clearly $\pi_n \geq 0$ for all $n \geq 0$ and $\sum_{n=0}^{\infty} \pi_n = 1$. Similarly, using the detailed balance equation, we obtain for any $n \geq 0$,

$$\lambda\pi_n - \pi_{n+1} = \lambda\pi_\infty^\sigma(n) - \pi_\infty^\sigma(n + 1) \geq \lambda\pi_\infty^\sigma(n, 1)\sigma(n, 1) - \pi_\infty^\sigma(n + 1) = 0.$$

Thus, to show feasibility of π, we must verify that (2.5a) and (2.5b) hold. To see this, observe that

$$
\begin{aligned}
\mathbf{E}^{\sigma}[h(X_{\infty},p)\mathbf{I}\{s=1\}] &= \sum_{n=0}^{\infty} \pi^{\sigma}_{\infty}(n,s=1)h(n,p) \\
&= \sum_{n=0}^{\infty} \pi^{\sigma}_{\infty}(n)\sigma(n,1)h(n,p) \\
&= \frac{1}{\lambda}\sum_{n=0}^{\infty} \pi^{\sigma}_{\infty}(n+1)h(n,p) \\
&= \frac{1}{\lambda}\sum_{n=0}^{\infty} \pi_{n+1}h(n,p) \\
&= \frac{1}{\lambda}\sum_{n=1}^{\infty} \pi_{n}h(n-1,p). \qquad \text{(A3)}
\end{aligned}
$$

Here, the third equality follows from the detailed balance condition. Since σ is feasible for (2.4), we have $\mathbf{E}[h(X_{\infty},p)|s=1] \geq 0$. From this, we conclude that $\sum_{n=1}^{\infty}\pi_n h(n-1,p) \geq 0$, and hence (2.5a) holds. Similarly, we have

$$
\begin{aligned}
\mathbf{E}^{\sigma}[h(X_{\infty},p)\mathbf{I}\{s=0\}] &= \sum_{n=0}^{\infty} \pi^{\sigma}_{\infty}(n,s=0)h(n,p) \\
&= \sum_{n=0}^{\infty} \pi^{\sigma}_{\infty}(n)(1-\sigma(n,1))h(n,p) \\
&= \sum_{n=0}^{\infty} \left(\pi^{\sigma}_{\infty}(n) - \frac{1}{\lambda}\pi^{\sigma}_{\infty}(n+1)\right) h(n,p) \\
&= \frac{1}{\lambda}\sum_{n=0}^{\infty} \left(\lambda\pi_n - \pi_{n+1}\right) h(n,p). \qquad \text{(A4)}
\end{aligned}
$$

Again, we have used the detailed balance condition in the third equality. Since σ is feasible for (2.4), we have $\mathbf{E}^{\sigma}[h(X_{\infty},p)|s=0] \leq 0$. From this and the preceding equalities, we conclude that (2.5b) holds. Finally, observe that

$$
\mathbf{E}^{\sigma}[\lambda\sigma(X_{\infty},1)] = \sum_{n=0}^{\infty} \lambda\pi^{\sigma}_{\infty}(n)\sigma(n,1) = \sum_{n=0}^{\infty} \pi^{\sigma}_{\infty}(n+1) = \sum_{n=1}^{\infty} \pi_n. \qquad \text{(A5)}
$$

Thus, we obtain that π is feasible for (2.5), with the same objective value as σ in (2.4).

Next, consider any feasible solution $\pi = \{\pi_n : n \geq 0\}$ for (2.5). We show that there exists a signaling mechanism σ feasible for (2.4) that attains the same objective value as π. Define

$\sigma : \mathbb{N}_0 \times \{0, 1\} \to [0, 1]$ as

$$\sigma(n, 1) = \begin{cases} \frac{\pi_{n+1}}{\lambda \pi_n} & \text{if } \pi_n > 0; \\ 0 & \text{otherwise.} \end{cases}$$

In order to verify that the obedience constraints hold for σ, we first compute the steady state distribution π^σ_∞ when all customers follow the obedient strategy. Using (A1) and (A2), we get $\pi^\sigma_\infty(n) = \pi_n$ for all $n \geq 0$. Thus, from (A3) and (A4) and from the fact that π is feasible for (2.5), we obtain that $\mathbf{E}^\sigma[h(X_\infty, p)\mathbf{I}\{s = 1\}] \geq 0$ and $\mathbf{E}^\sigma[h(X_\infty, p)\mathbf{I}\{s = 0\}] \leq 0$. After conditioning on the appropriate event, we obtain that σ satisfies the obedience constraints. Finally, using (A5), we conclude that σ achieves the same objective value in (2.4) as π in the linear program (2.5). □

The following lemma is used in the proof of Theorem 2.1 to show that the maximum in the linear program (2.5) is attained.

Lemma A.1. *Let $\mathcal{D}$ denote the set of all feasible solutions $\{\pi_n : n \geq 0\}$ to (2.5) of the following form: there exists an $N \geq M_p$, such that $\pi_n = \lambda \pi_{n-1}$ for all $n < N$, $0 < \pi_N \leq \lambda \pi_{N-1}$ and $\pi_n = 0$ for $n > N$. (Note N can equal ∞.) Then, the set $\mathcal{D}$ is compact under the weak topology.*

Proof. We will show that the set of distributions $\mathcal{D}$ is tight. The result then follows from Prokhorov's theorem. To show tightness, we prove that for any $\epsilon > 0$, there exists an N such that for any $\pi \in \mathcal{D}$, we have $\sum_{n>N} \pi_n < \epsilon$.

Fix an $\epsilon > 0$. First note that if $\lambda < 1$, then for any feasible solution π, we have $\pi_n \leq \lambda^n \pi_0 \leq \lambda^n$. Hence, we obtain that for all large enough N, $\sum_{n>N} \pi_n \leq \sum_{n>N} \lambda^n < \lambda^N/(1 - \lambda) < \epsilon$. Thus, for the rest of the proof, suppose $\lambda \geq 1$.

Let N be the first positive integer such that $\sum_{n=1}^{N-1} \lambda^n h(n - 1, p) < 0$ (since $h(n, p) \leq h(M_p, p) < 0$ for all $n \geq M_p$, there must be such a value of N). Note that for $\lambda \geq 1$ and for

$\pi \in \mathcal{D}$, there must be an $m < \infty$ such that $\pi_m < \lambda \pi_{m-1}$. For a feasible $\pi \in \mathcal{D}$, let L be such that $\pi_n = \lambda^n \pi_0$ for all $n < L$, $0 \leq \pi_L < \lambda \pi_{L-1}$ and $\pi_n = 0$ for all $n > L$. If $L > N$, we have

$$\sum_{n=1}^{\infty} \pi_n h(n-1,p) = \pi_0 \sum_{n=1}^{L-1} \lambda^n h(n-1,p) + h(L-1,p)\pi_L \leq \pi_0 \sum_{n=1}^{N-1} \lambda^n h(n-1,p) < 0,$$

contradicting the fact that $\pi \in \mathcal{D}$. Thus, for all $\pi \in \mathcal{D}$, we have $L \leq N$. and hence $\sum_{n>N} \pi_n = 0 < \epsilon$. Thus, by Prokhorov's theorem, the set $\mathcal{D}$ is compact. □

The following simple lemma states that the throughput is increasing in the threshold, and is used in the proof of Theorem 2.3.

Lemma A.2. *The throughput of the threshold signaling mechanism σ^x is monotonically increasing in $x \in \mathbb{R}_+$.*

Proof. Note that for $x = N + q$ with $N \in \mathbb{N}_0$ and $q \in [0,1]$, we have $\pi_\infty^x(0) = \left(\sum_{i=0}^{N} \lambda^i + \lambda^{N+1} q\right)^{-1}$, where $\{\pi_\infty^x(n) : n \geq 0\}$ denotes the steady state distribution of the queue under the signaling mechanism σ^x. This follows from the fact that under σ^x, there are at most $N+1$ customers in the queue, with a new customer joining the queue with probability 1 if the number of customers already in the queue is strictly less than N, and joining with probability q if the number of customers is equal to N, and balking otherwise. Thus, $\pi_\infty^x(0)$ is strictly decreasing in $x = N + q$. The result then follows from the fact that throughput satisfies $\mathsf{Th}(\sigma^x) = \lambda(1 - \pi_\infty^x(0))$. □

A.1.1 Proof of Theorem 2.3

We now present the proof of Theorem 2.3, obtaining analytical expression for the optimal threshold in the case of linear utility.

Proof of Theorem 2.3. Consider a threshold mechanism σ^x with $x \geq M_p$. We seek to find the largest value of x for which the obedient strategy is a customer equilibrium. Then, by Lemma A.2 and Theorem 2.1, we obtain the threshold mechanism σ^x is optimal.

To begin, note that since $x \geq M_p$, if a customer observes a signal $s = 0$, then the number of customers in the queue is at least $\lfloor x \rfloor = M_p$, and hence the expected payoff on joining the queue is at most $h(M_p, p) \leq 0$. Hence, leaving on seeing signal $s = 0$ is indeed optimal.

We have for $x = N + q \geq M_p$,

$$\pi_\infty^x(n|s=1) = \frac{\lambda^n \mathbf{I}\{n < N\} + q\lambda^N \mathbf{I}\{n = N\}}{\sum_{k<N} \lambda^k + q\lambda^N}$$

This implies,

$$\mathbf{E}^{\sigma^x}[h(X_\infty, p)|s=1] = \frac{\sum_{k<N} \lambda^k (1 - p - c(k+1)) + q\lambda^N (1 - p - c(N+1))}{\sum_{k<N} \lambda^k + q\lambda^N}$$

Thus, for joining the queue to be optimal for a customer on seeing a signal $s = 1$, we must have

$$\sum_{k<N} \lambda^k (1 - p - c(k+1)) + q\lambda^N (1 - p - c(N+1)) \geq 0. \tag{A6}$$

We consider the following two cases separately:

Case 1: $\lambda = 1$**.** In this case, the equation (A6) becomes

$$(1 - p - c)N - \frac{c}{2}N(N-1) + q(1 - p - c(N+1)) \geq 0. \tag{A7}$$

We first consider the case where $q = 0$ to find the largest N that satisfies this equation. The largest such value of N is

$$N^* = \left\lfloor \frac{2(1-p)}{c} - 1 \right\rfloor.$$

Unless the expression inside the floor-operator on the right hand side is an integer, we have $(1 - p - c)N - \frac{c}{2}N(N-1) > 0$, implying we can set $q > 0$ and not violate (A7). The largest value of $q = q^*$ that can be set is for which (A7) is an equality. (Note that q^* cannot be equal to 1, by definition of N^*.) From this, we obtain the following expression for q^*:

$$q^* = \frac{(1-p-c)N^* - \frac{c}{2}N^*(N^*-1)}{c(N^*+1)+p-1}$$

Case 2: $\lambda \neq 1$. We, again, first consider the case where $q = 0$ and and seek the largest value of N that satisfies (A6). For any value N that satisfies (A6), upon adding up the summations, we obtain

$$(1-p)\frac{1-\lambda^N}{1-\lambda} - c\left(\frac{1-\lambda^N}{1-\lambda} + \frac{\lambda - N\lambda^N + (N-1)\lambda^{N+1}}{(1-\lambda)^2}\right) \geq 0,$$

which on simplifying yields,

$$\frac{(1-p-c)(1-\lambda)}{\lambda c}(1-\lambda^N) \geq 1 - N\lambda^{N-1} + (N-1)\lambda^N.$$

Let $\alpha = \frac{(1-p-c)(1-\lambda)}{\lambda c}$ and $\beta = \frac{1-\lambda}{\lambda}$. Then, we obtain after some algebra,

$$(N\beta + 1 - \alpha)\,\lambda^N \geq 1 - \alpha. \tag{A8}$$

Note that if $\alpha \geq 1$, implying that $\lambda \leq 1 - \frac{c}{1-p}$, then the right-hand side is non-negative for all $N \geq 1$. Thus, all values of $N \geq M_p$ satisfy this equation, and hence we obtain $N^* = \infty$. In other words, the optimal signaling mechanism always signals the customer to join the queue.

Suppose now that $\alpha < 1$. Then, multiplying both sides of (A8) by $\left(\lambda^{\frac{1}{\beta}}\right)^{1-\alpha} > 0$ gives us

$$(N\beta + 1 - \alpha)\left(\lambda^{\frac{1}{\beta}}\right)^{N\beta+1-\alpha} \geq (1-\alpha)\left(\lambda^{\frac{1}{\beta}}\right)^{1-\alpha}.$$

Let $\psi = N\beta + 1 - \alpha$ and $\gamma = \lambda^{1/\beta}$. Note that for all $\lambda \neq 1$, we have $\gamma < 1$. The preceding equation can be written as $(1-\alpha)\gamma^{1-\alpha} \leq \psi\gamma^{\psi}$. After multiplying both sides by $\log(1/\gamma) > 0$ and some algebra, we obtain

$$\psi \log\left(\frac{1}{\gamma}\right)\exp\left(-\psi\log\left(\frac{1}{\gamma}\right)\right) \geq (1-\alpha)\log\left(\frac{1}{\gamma}\right)\exp\left(-(1-\alpha)\log\left(\frac{1}{\gamma}\right)\right). \tag{A9}$$

For $x > 0$, let $H(x)$ be the function defined implicitly by $H(x)\exp(-H(x)) = x\exp(-x)$ with $H(x) \neq x$ for $x \neq 1$. Observe that if $x > 1$, then $H(x) < 1$ and if $x < 1$, then $H(x) > 1$, with $H(1) = 1$.

Note that if $\lambda > 1$, then $\beta < 0$, $\alpha < 0$, which implies $1 - \alpha \geq 1$. Further, we obtain that for $\lambda > 1$, $\gamma \leq e^{-1}$, which implies $\log(1/\gamma) \geq 1$. Hence $z = (1-\alpha)\log(1/\gamma) \geq 1$. On the other hand, if $1 - \frac{c}{1-p} < \lambda < 1$, then $\beta > 0$, $1 - \alpha \in [0, 1]$, and furthermore $\log(1/\gamma) \leq 1$. Hence, $z = (1-\alpha)\log(1/\gamma) \leq 1$. Using these facts, and the definition of $H(\cdot)$, we obtain from (A9),

$$\begin{aligned} H\left((1-\alpha)\log\left(\frac{1}{\gamma}\right)\right) \leq \psi \log\left(\frac{1}{\gamma}\right) \leq (1-\alpha)\log\left(\frac{1}{\gamma}\right), \qquad & \text{if } \lambda > 1; \\ (1-\alpha)\log\left(\frac{1}{\gamma}\right) \leq \psi \log\left(\frac{1}{\gamma}\right) \leq H\left((1-\alpha)\log\left(\frac{1}{\gamma}\right)\right), \qquad & \text{if } \lambda < 1. \end{aligned}$$

Using the fact that $\psi = N\beta + 1 - \alpha$, and noting that $\beta < 0$ if $\lambda > 1$ and $\beta > 0$ if $\lambda < 1$, we get

$$N \leq \frac{1}{\beta \log\left(\frac{1}{\gamma}\right)} \left(H\left((1-\alpha)\log\left(\frac{1}{\gamma}\right)\right) - (1-\alpha)\log\left(\frac{1}{\gamma}\right)\right).$$

Since N^* is the largest such value of N, we have

$$N^* = \left\lfloor \frac{1}{\beta \log\left(\frac{1}{\gamma}\right)} \left(H\left((1-\alpha)\log\left(\frac{1}{\gamma}\right)\right) - (1-\alpha)\log\left(\frac{1}{\gamma}\right)\right) \right\rfloor.$$

Using the definition of the Lambert-W function and its two branches W_0 and W_{-1} (see Borgs et al. (2014)), it can be shown that for $x > 0$, we have $H(x) = -W_i(xe^{-x})$, where $i = 0$ if $x > 1$ and $i = -1$ if $x < 1$. Upon letting $\kappa = (1-\alpha)\log\left(\frac{1}{\gamma}\right) = (\frac{1-p}{c} - \frac{1}{1-\lambda})\log(\lambda)$, we obtain

$$N^* = \left\lfloor \frac{1}{\log(\lambda)} \left(W_i\left(-\kappa e^{-\kappa}\right) + \kappa\right) \right\rfloor,$$

where $i = 0$ if $\lambda > 1$ and $i = -1$ if $1 - \frac{c}{1-p} < \lambda < 1$.

Finally, observe that unless the expression inside the floor-operator in the expression for N^* is an integer, we have $\sum_{k<N^*} \lambda^k (1 - c(k+1)) > 0$, and we can set $q > 0$ without violating (A6). The largest value of $q = q^*$ that can be set is for which (A6) is an equality. (Note that q^* cannot be equal to 1, by definition of N^*.) From this, we obtain

$$q^* = \frac{\sum_{k<N^*} \lambda^k (1 - p - c(k+1))}{\lambda^{N^*}(c(N^*+1) + p - 1)}.$$

This completes the proof. □

A.1.2 Proof of Theorem 2.5

In this subsection, we prove Theorem 2.5. The proof follows a similar structure to that of Theorem 2.2: we use the structure of the optimal state-and-type-dependent pricing mechanism in a fully observable queue to construct a threshold signaling mechanism (with fixed prices) that attains the same revenue. As the first step, we analyze the optimal state-and-type-dependent prices in a fully observable queue, where the service provider sets a price $p_i(n)$ for customer type i and queue-length n. We have the following lemma.

Lemma A.3. *The optimal state-and-type-dependent pricing mechanism in a fully observable queue satisfies*

$$p_i(n) = \begin{cases} u_i(n) & \text{if } n < \kappa_i; \\ \infty & \text{if } n \geq \kappa_i, \end{cases} \tag{A10}$$

for some $\kappa = (\kappa_1, \ldots, \kappa_K) \in \mathbb{N}_0^K$. Furthermore, this mechanism is welfare-optimal.

Proof of Lemma A.3. We begin by formulating the pricing problem in a fully observable queue as an infinite-horizon Markov decision process (MDP) with average rewards. We consider the embedded discrete time chain with states as follows. For each $n \geq 0$ and $i \in \{1, \ldots, K\}$, let (n, i) denote the state where there are n customers already in the queue and a customer of type i has arrived. Similarly, for $n \geq 0$, let $(n, 0)$ denote the state after a customer has departed leaving n customers in the queue. Note that the service provider must choose a price $p_i(n) \geq 0$ at state (n, i) for $n \geq 0$ and $i \in \{1, \ldots, K\}$. (At state $(n, 0)$, the service provider chooses a dummy action.)

First, note that since $\lim_{X\to\infty} u_i(X) < 0$ for all i, there exists an N such that the $u_i(n) < 0$ for all i and $n > N$. Since $p_i(n) > u_i(n)$ for all $n > N$ and each i, no customer will join the queue when there are at least N customers already in the queue. Thus, it follows that the MDP is unichain (Puterman 1994) and the optimal prices can be found by solving the Bellman equation. Let V denote the average revenue under the optimal pricing mechanism, and let $g(n,i)$ denote the *bias* (Puterman 1994) of each state (n,i).

The Bellman equation for the pricing problem can then be written as follows: for each $n \geq 0$ and $i \in \{1,\ldots,K\}$, we have

$$\begin{aligned} V + g(n,i) = \max_{p_i(n)\geq 0} \Bigg[& \mathbf{I}\{p_i(n) \leq u_i(n)\} \left(p_i(n) + \sum_{j=1}^{K} \frac{\lambda_j}{1+\Lambda} g(n+1,j) + \frac{1}{1+\Lambda} g(n,0) \right) \\ & + \mathbf{I}\{p_i(n) > u_i(n)\} \left(\sum_{j=1}^{K} \frac{\lambda_j}{1_n+\Lambda} g(n,j) + \frac{1_n}{1_n+\Lambda} g(n-1,0) \right) \Bigg], \end{aligned} \tag{A11}$$

and

$$V + g(n,0) = \sum_{j=1}^{K} \frac{\lambda_j}{1_n+\Lambda} g(n,j) + \frac{1_n}{1_n+\Lambda} g(n-1,0). \tag{A12}$$

Here, we define $1_n \triangleq \mathbf{I}\{n > 0\}$ and recall that $\Lambda = \sum_{j=1}^{K} \lambda_j$. The first equation follows from the fact that if $p_i(n) \leq u_i(n)$, a customer of type i will join the queue at state (n,i), yielding an immediate revenue of $p_i(n)$. Subsequently, the queue state transitions to $(n+1,j)$ with probability $\lambda_j/(1+\Lambda)$ for $j \in \{1,\ldots,K\}$, and to state $(n,0)$ with probability $1/(1+\Lambda)$. On the other hand, if $p_i(n) > u_i(n)$, then the customer does not join the queue at state (n,i), yielding no immediate revenue and similar subsequent transitions. The second equation follows from the fact that the service provider has a single dummy action at state $(n,0)$ that yields no immediate revenue.

From the Bellman equation, it follows that one can always restrict to $p_i(n) \in \{u_i(n), \infty\}$: the price $p_i(n) = \infty$ performs equally as well as any $p_i(n) > u_i(n)$, and any $p_i(n) < u_i(n)$ is strictly dominated by $p_i(n) = u_i(n)$. Using this, we can write (A11) as

$$\begin{aligned} V + g(n,i) = \max \Bigg[& u_i(n) + \sum_{j=1}^{K} \frac{\lambda_j}{1+\Lambda} g(n+1,j) + \frac{1}{1+\Lambda} g(n,0)\ , \\ & \sum_{j=1}^{K} \frac{\lambda_j}{1_n+\Lambda} g(n,j) + \frac{1_n}{1_n+\Lambda} g(n-1,0) \Bigg], \end{aligned} \tag{A13}$$

where the optimal price is $u_i(n)$ if the first term attains the maximum, and ∞ otherwise. Thus, it remains to show that the optimal pricing mechanism has a threshold structure, i.e., $p_i(n) < \infty$ implies $p_i(m) < \infty$ for all $m < n$.

Substituting (A12) into (A13) and after simplifying, we obtain

$$g(n,i) = \max\{u_i(n) + g(n+1,0), g(n,0)\} = g(n,0) + (u_i(n) - \eta(n))^+,$$

where $\eta(n) \triangleq g(n,0) - g(n+1,0)$ for all n. Substituting this expression back into (A12) and simplifying, we obtain the following equation that holds for all $n \geq 0$:

$$V = \sum_{i=1}^{K} \frac{\lambda_i}{1_n + \Lambda} (u_i(n) - \eta(n))^+ + \frac{1_n}{1_n + \Lambda} \eta(n-1). \tag{A14}$$

Note that $p_i(n) < \infty$ if and only if $u_i(n) \geq \eta(n)$. Thus, to show that the optimal prices have a threshold structure, we must show that if $u_i(n) \geq \eta(n)$ for some n, then $u_i(m) \geq \eta(m)$ for all $m < n$. Since $u_i(n)$ is non-increasing in n for each i, it suffices to show that $\eta(n)$ is non-decreasing in n. We prove this latter statement by induction.

First, note that since $\lim_{X\to\infty} u_i(X) < 0$ for all $i \in \{1, \ldots, K\}$, there exists an $N > 0$ such that $u_i(n) < 0$ for each i and $n > N$. By our earlier argument, this implies that optimal prices satisfy $p_i(n) = \infty$ for all i and $n > N$, which in turn implies $u_i(n) < \eta(n)$ for all i and $n > N$. Then, using (A14), we obtain $\eta(n) = V(1+\Lambda)$ for all $n \geq N$. Hence, $\eta(n) \geq \eta(n-1)$ for all $n > N$.

Now suppose $\eta(n) \geq \eta(n-1)$ for some $n \geq 2$. Note that since $u_i(n) \leq u_i(n-1)$, this implies that $(u_i(n) - \eta(n))^+ \leq (u_i(n-1) - \eta(n-1))^+$. From (A14) and using the fact that $n \geq 2$, we obtain

$$\begin{aligned}
V &= \sum_{i=1}^{K} \frac{\lambda_i}{1+\Lambda} (u_i(n) - \eta(n))^+ + \frac{1}{1+\Lambda} \eta(n-1) \\
&\leq \sum_{i=1}^{K} \frac{\lambda_i}{1+\Lambda} (u_i(n-1) - \eta(n-1))^+ + \frac{1}{1+\Lambda} \eta(n-1) \\
&= V - \frac{1}{1+\Lambda} \eta(n-2) + \frac{1}{1+\Lambda} \eta(n-1).
\end{aligned}$$

Thus, we have $\eta(n-1) \geq \eta(n-2)$. This completes the induction step, and we conclude that $\eta(n)$ is non-decreasing in n. Thus, if $u_i(n) \geq \eta(n)$ for some n, then $u_i(m) \geq \eta(m)$ for all $m < n$, and hence the optimal prices have a threshold structure.

Finally, it is straightforward to show that the problem of welfare optimization (where the service provider performs admission control to maximize social welfare) can be written as a dynamic program with the same Bellman equation, given by (A12) and (A13). This implies that the optimal state-and-type-dependent pricing mechanisms is also welfare optimal. $\square$

We conclude with the proof of Theorem 2.5.

Proof of Theorem 2.5. From Lemma A.3, let $\kappa = (\kappa_1, \cdots, \kappa_K)$ denote the thresholds in the optimal state-and-type-dependent pricing mechanism. Let X_i^κ denote the steady state distribution of the queue under this pricing mechanism.

For the unobservable queue, consider the signaling mechanism σ, where $\sigma(n, i, 1) = \mathbf{I}\{n < \kappa_i\}$ for each $i \in \{1, \cdots, K\}$, and the fixed (type-dependent) prices $p_i = \mathbf{E}\left[u(X_\infty^\kappa) | X_\infty^\kappa < \kappa_i\right]$. Using the same argument as in the proof of Theorem 2.2, it is straightforward to show that, for this setting, obedience is an equilibrium, and that under the obedient equilibrium, the service provider's revenue is same as that of the optimal state-and-type-dependent pricing mechanism.

Finally, since the latter mechanism is welfare-optimal and has zero customer surplus, we conclude that the mechanism σ and the prices p_i together constitute the optimal fixed price and signaling mechanism. $\square$

A.2 Comparison of the fully-revealing and the no-information mechanisms

In this section, we briefly compare the fully-revealing and no-information mechanisms in the case of linear expected utility $u(X) = 1 - c(X+1)$ and a fixed price p. Observe that for $p > 1 - c$, no customer will join the queue to obtain service, and hence the service provider's revenue is zero. We restrict our attention to $p \in [0, 1-c]$.

For the fully revealing mechanism, we find the throughput to be

$$\mathsf{Th}^{\mathsf{full}} = \lambda \left(\frac{1 - \lambda^{M_p}}{1 - \lambda^{M_p+1}} \right),$$

where $M_p = \left\lceil \frac{1-c-p}{c} \right\rceil$. The optimal revenue for the full-information signal is given by

$$R^{\mathsf{full}} = \max_p \left(\frac{\lambda - \lambda^{M_p+1}}{1 - \lambda^{M_p+1}} \right) p.$$

In the case of the no-information mechanism, a customer strategy is a probability q with which a customer joins the queue. We can view the queue as a thinned M/M/1 queue with arrival rate $q\lambda$. Recall that the stationary distribution for such a queue is $\frac{q\lambda}{1-q\lambda}$.

Note that $q = 0$ is not an equilibrium for $p < 1 - c$: if so, joining the queue would have utility $1 - c - p > 0$. We see that $q = 1$ is an equilibrium if and only if the utility for joining the queue $(1 - p - c(\frac{\lambda}{1-\lambda} + 1))$ is at least that of not (0), or, equivalently, if $\lambda \leq 1 - \frac{c}{1-p}$. Otherwise, if $\lambda > 1 - \frac{c}{1-p}$, we must have a mixed strategy equilibrium $q \in (0,1)$. For this to be an equilibrium, the utility for joining the queue $(1 - p - c(\frac{q\lambda}{1-q\lambda} + 1))$ must equal the utility for not joining the queue (0), so that a mixed strategy is optimal. This is equivalent to $q = \frac{1}{\lambda}\left(1 - \frac{c}{1-p}\right)$. Putting these cases together, we get that for any $p \in [0, 1-c]$, the customer equilibrium f is given by $q = \min\{\frac{1}{\lambda}(1 - \frac{c}{1-p}), 1\}$, with the corresponding throughput given by $\mathsf{Th}^{\mathsf{no-info}} = \min\{1 - \frac{c}{1-p}, \lambda\}$. Maximizing the revenue $p \cdot \mathsf{Th}^{\mathsf{no-info}}$ over values of $p \in [0, 1-c]$, we obtain the optimal price p^* to be

$$p^* = \begin{cases} 1 - \frac{c}{1-\lambda} & \text{if } \lambda \leq 1 - \sqrt{c}; \\ 1 - \sqrt{c} & \text{otherwise,} \end{cases}$$

with corresponding revenue given by

$$R^{\textsf{no-info}} = \begin{cases} \lambda - \frac{c\lambda}{1-\lambda} & \text{if } \lambda \leq 1 - \sqrt{c}; \\ (1 - \sqrt{c})^2 & \text{otherwise.} \end{cases}$$

From the preceding discussion, we observe that for values where $\lambda < 1 - \frac{c}{1-p}$, we have $\textsf{Th}^{\textsf{full}} < \textsf{Th}^{\textsf{no-info}}$, implying that sharing no information about the queue with customers achieves higher throughput than revealing the number of customers in the queue. On the other hand, observe that when $p = 1 - c - \epsilon$ for small enough $\epsilon > 0$, we have $M_p = 1$, and hence,

$$\lim_{\lambda \to \infty} R^{\textsf{full}} \geq \lim_{\lambda \to \infty} (1 - c - \epsilon) \left(\frac{\lambda - \lambda^2}{1 - \lambda^2} \right) = 1 - c - \epsilon.$$

However, since $\sqrt{c} > c + \epsilon$ for small enough c, we have

$$\lim_{\lambda \to \infty} R^{\textsf{no-info}} = (1 - \sqrt{c})^2 = 1 - 2\sqrt{c} + c < 1 - c - \epsilon.$$

Thus, in this limiting regime, we have that revealing the number of customers in the queue obtains a higher revenue than not revealing, as seen for large values of λ in Figure 2.1c.

APPENDIX B

APPENDIX TO CHAPTER 3

B.1 Proofs from Sections 3.3 and 3.4

Proof of Lemma 3.1. We consider the following abstraction: there are $N_{1,\max}$ agents (each with the same prior as the firm's) from which a set of N_1 customers is drawn independently and uniformly. Intuitively, a customer has one piece of information to update her prior, that she is present at time 1, which has a higher likelihood when N_1 is larger.

To see this concretely, fix a focal agent and let B be the event that the focal agent is present at time 1. By Bayes rule,

$$\begin{aligned}\Phi_C(q,n_1,n_2) &= \mathbf{P}(Q=q,N_1=n_1,N_2=n_2|B)\\ &= \frac{\mathbf{P}(B|Q=q,N_1=n_1,N_2=n_2)\,\mathbf{P}(Q=q,N_1=n_1,N_2=n_2)}{\sum_{\hat{q},\hat{n}_1,\hat{n}_2}\mathbf{P}(B|Q=\hat{q},N_1=\hat{n}_1,N_2=\hat{n}_2)\,\mathbf{P}(Q=\hat{q},N_1=\hat{n}_1,N_2=\hat{n}_2)}\end{aligned} \tag{A1}$$

By definition, $\mathbf{P}(Q=q,N_1=n_1,N_2=n_2)=\Phi(q,n_1,n_2)$. Also, if there are n_1 customers present, the probability the focal agent is one of them is $\frac{n_1}{N_{1,\max}}$, so $\mathbf{P}(B|Q=q,N_1=n_1,N_2=n_2)=\frac{n_1}{N_{1,\max}}$. Substituting these values in (A1) and canceling the common factor $N_{1,\max}$ yields the lemma statement. □

Proof of Lemma 3.2. Consider a signaling mechanism $\mathcal{S}=(S,\sigma)$ and customer equilibrium $\mathbf{f}$. We prove this lemma by first showing we can find a signaling mechanism $\mathcal{U}=(U,\upsilon)$ and customer equilibrium $\mathbf{g}$ that achieves the same revenue where $\mathbf{g}$ is pure. Then we will construct a signaling mechanism where, further, the signals are binary. Finally, we will construct a signaling mechanism that is additionally symmetric.

Given a signaling mechanism $\mathcal{S}=(S,\sigma)$ and customer equilibrium $\mathbf{f}$, consider a new signaling mechanism $\mathcal{U}=(U,\upsilon)$, where $U=S\times\{0,1\}$. Any $\mathbf{t}\in U^{N_1}$ can be represented as

$\{\mathbf{s}, \mathbf{r}\}$ where $\mathbf{s} \in S^{N_1}$ and $\mathbf{r} \in \{0,1\}^{N_1}$. Then, σ_1 is given by

$$\sigma_1(Q, N_1, \{\mathbf{s}, \mathbf{r}\}) = \sigma(Q, N_1, \mathbf{s}) \prod_{i=0}^{N_1} \left(I(\mathbf{r}_i = 0)(1 - f_i(s_i)) + I(\mathbf{r}_i = 1) f_i(s_i) \right).$$

Notice that this signaling mechanism is identical to $\mathcal{S}$, but to each customer receiving a signal s, we are also providing the results of an independent coin flip with odds $f_i(s)$. Now consider the strategy profile $\mathbf{g}$ under the signaling mechanism $\mathcal{U}$, where $g_i(s,1) = 1$ and $g_i(s,0) = 0$. We will show $\mathbf{g}$ is an customer equilibrium.

First, we will consider the utility of customer i for each action. Assume all other customers behave as they would under $\mathcal{S}$, so that the expected utility for each action is the same as before. If our customer is given the signal $(s,1)$, then $f_i(s) > 0$. Thus, it must be the case that $\mathbf{E}_C[h(Q, N_1, \widehat{D}_{-i}) | s_i = s] \geq 0$, and it is optimal for customer i to buy now. If, instead, customer i received $(s,0)$ then $f_i(s) < 1$ and $\mathbf{E}_C[h(Q, N_1, \widehat{D}_{-i}) | s_i = s] \leq 0$, so it is optimal for customer i to wait. Hence, it is optimal for our customer to follow strategy g_i and $\mathbf{g}$ is an equilibrium to $\mathcal{U}$.

Given the partial signal s, the probability of customer i choosing to buy now is equal to what it was under $\mathcal{S}$. Since the partial signals s are sent according to the same mechanism, $\mathcal{S}$ and $\mathcal{U}$ achieve the same expected revenue. Also, g_i is pure for all i, as desired.

Next, consider a signaling mechanism $\mathcal{S} = (S, \sigma)$ and customer equilibrium $\mathbf{f}$ where $\mathbf{f}$ is pure. Let $S_{i,j} = \{s \in S | f_i(s) = j\}$ for $j = 0, 1$ and $i \leq N_1$. We define a new signaling mechanism $\mathcal{U} = (U, \upsilon)$ where $U = \{0, 1\}$ and for any $\mathbf{t} \in \{0,1\}^{N_1}$,

$$\upsilon(Q, N_1, \mathbf{t}) = \sum_{\mathbf{s} \in S^{N_1}} I(\mathbf{s}_j \in S_{j, \mathbf{t}_j} \text{ for all } j) \sigma(Q, N_1, \mathbf{s}).$$

Notice that upon receiving the signal j, a customer knows their signal under $\mathcal{S}$ was in $S_{i,j}$, and hence their optimal action is taking action i as they would have under $\mathcal{S}$, assuming all other players make the same decision they did before. Hence, $\mathbf{g}$ is an equilibrium. By how we defined υ, it achieves the same revenue.

Finally, consider a signaling mechanism $\mathcal{S} = (\{0,1\}, \sigma)$ and customer equilibrium $\mathbf{f}$ where $\mathbf{f}$ is pure. We construct a new signaling mechanism, $\mathcal{U} = (\{0,1\}, \upsilon)$, that does the following: First it shuffles the labels on customers, then it uses the signaling mechanism $\mathcal{S}$. Let $P(\mathbf{t})$ be the set of permutations of the tuple $\mathbf{t}$. Then υ is

$$\upsilon(Q, N_1, \mathbf{t}) = \frac{\sum_{s \in P(\mathbf{t})} \sigma(Q, N_1, \mathbf{t})}{|P(\mathbf{t})|}.$$

It follows that $\upsilon(Q, N_1, \mathbf{t}) \geq 0$ and $\sum_{\mathbf{t} \in \{0,1\}^{N_1}} \upsilon(Q, N_1, \mathbf{t}) = 1$. Next, we see that upon receiving signal j, it is optimal for customer i to choose action j: no matter which customer i was relabeled to, it was optimal for that customer to follow action j. Hence, g is an equilibria.

Finally, note that the probability of asking D total customers to join (for any D) is the same in both settings: we simply shuffled the probability of sending signals with the same D. Hence, the expected revenue is identical. □

Proof of Lemma 3.3. If $q < n$, we see that $h(q, n, n-1) = (v - p_1)\frac{q}{n} + c > 0$. If instead $q \geq n$, then

$$\begin{aligned} h(q, n, d-1) &= (v - p_1) + c - \mathbf{E}\left[\min\left\{\frac{Q - d + 1}{N_1 + N_2 - d + 1}, 1\right\}(v - p_2)\middle| Q = q, N_1 = n\right] \\ &\leq (v - p_1) + c - \mathbf{E}\left[\min\left\{\frac{Q - d}{N_1 + N_2 - d}, 1\right\}(v - p_2)\middle| Q = q, N_1 = n\right] = h(q, n, d) \end{aligned}$$

and thus $h(q, n, d-1) \geq 0$ implies $h(q, n, n-1) \geq 0$, as desired. □

Proof of Lemma 3.4. We first assume $h(q, n, n-1) \geq 0$. If $h(q, n, d-1) \leq 0$, then Lemma (3.4.1) holds. If not, suppose $h(q, n, d-1) \geq 0$. We would like to show that $dh(q, n, d-1) \leq (d+1)h(q, n, d)$ holds for all d where $h(q, n, d-1) \geq 0$. This would imply that $dh(q, n, d-1) \leq nh(q, n, n-1)$, as desired.

Now, suppose $q > d$. Then, like in the proof of Lemma 3.3,

$$h(q,n,d-1) = (v-p_1) + c - \mathbf{E}\left[\min\left\{\frac{Q-d+1}{N_1+N_2-d+1}, 1\right\}(v-p_2)\middle| Q=q, N_1=n\right]$$
$$\leq (v-p_1) + c - \mathbf{E}\left[\min\left\{\frac{Q-d}{N_1+N_2-d}, 1\right\}(v-p_2)\middle| Q=q, N_1=n\right] = h(q,n,d).$$

Hence, $dh(q,n,d-1) \leq (d+1)h(q,n,d)$. If $d \geq q$, then

$$dh(q,n,d-1) = d\left((v-p_1)\frac{q}{d} + c\right) \leq (d+1)\left((v-p_1)\frac{q}{d+1} + c\right) = (d+1)h(q,n,d),$$

Thus $dh(q,n,d-1) \leq (d+1)h(q,n,d)$ and $dh(q,n,d-1) \leq nh(q,n,n-1)$, as desired.

Now suppose $h(q,n,n-1) < 0$. If $q < n$, we see that $h(q,n,n-1) = (v-p_1)\frac{q}{n} + c > 0$, which is a contradiction. Thus $q \geq n$. For any d, we have

$$h(q,n,d-1) = (v-p_1) + c - \mathbf{E}\left[\min\left\{\frac{Q-d+1}{N_1+N_2-d+1}, 1\right\}(v-p_2))\middle| Q=q, N_1=n\right]$$
$$\leq (v-p_1) + c - \mathbf{E}\left[\min\left\{\frac{Q-d}{N_1+N_2-d}, 1\right\}(v-p_2)\middle| Q=q, N_1=n\right] = h(q,n,d).$$

which implies $h(q,n,d-1) \leq h(q,n,n-1)$.

Thus, $\frac{r(q,n,d)-r(q,n,0)}{r(q,n,n)-r(q,n,0)} = \frac{d}{n}$ and

$$dh(q,n,d-1) \leq \left(\frac{r(q,n,d)-r(q,n,0)}{r(q,n,n)-r(q,n,0)}\right) nh(q,n,n-1)$$

and hence Lemma (3.4.2) holds. □

Proof of Theorem 3.1. Consider any solution, π, to (3.3). In this proof, we construct a new solution, $\hat{\pi}$ that has a weakly higher payoff for the seller and $\hat{\pi}(q,n,d) = 0$ whenever (and hence $\hat{\sigma}(\hat{q},\hat{n},\hat{d}) = 0$) whenever $h(q,n,n-1) \geq 0$. Consider the following solution $\hat{\pi}$:

$$\hat{\pi}(q,n,d) = \begin{cases} 0 & h(q,n,n-1) \geq 0, d \neq n \\ \Phi(q,n) & h(q,n,n-1) \geq 0, d = n \\ \pi(q,n,d) & \text{else} \end{cases}$$

It follows that this solution satisfies (3.3d) and (3.3c). We consider (3.3a): Consider any q, n such that $h(q,n,n-1) \geq 0$. By Lemma (3.4.1), $dh(q,n,n-1) \leq nh(q,n,d-1)$ for all $d \leq n$. Since $\hat{\pi}(q,n,n) = \sum_{d=0}^{n} \pi(q,n,d)$, the change to the left-hand-side of (3.3a) is $\sum_{(q,n)\in\Theta} \sum_{d=1}^{n} \pi(q,n,d)(nh(q,n,n-1) - dh(q,n,n-1)) \geq 0$, and so (3.3a) is satisfied by $\hat{\pi}$. We next consider (3.3b): We claim each nonzero term in the constraints sum is negative. If for some n, q, d, $(n-d)h(q,n,d)\hat{\pi}(q,n,d) > 0$, then $d < n$ and $h(q,n,d) \geq 0$. However, if $h(q,n,d) \geq 0$, then by Lemma 3.3, $h(q,n,n-1) \geq 0$ and $\hat{\pi}(q,n,d) = 0$. Hence, (3.3b) is satisfied by $\hat{\pi}$. The change to the objective is $\sum_{(q,n)\in\Theta} \sum_{d=0}^{n} \pi(q,n,d)(r(q,n,n) - r(q,n,n)) \geq 0$, and $\hat{\pi}$ is a weakly better solution.

We restrict our attention to π where this transformation is applied. So if $\pi(q,n,d) > 0$ and $h(q,n,n-1) \geq 0$, then $d = n$. The constraint (3.3b) is non-binding, and we can restrict our attention to the remaining constraints.

Next, suppose $h(\hat{q},\hat{n},\hat{n}-1) < 0$, which implies $h(\hat{q},\hat{n},d) < 0$ for all $d < \hat{n}-1$ by Lemma 3.3. Let $\gamma = \frac{r(\hat{q},\hat{n},\hat{d}) - r(\hat{q},\hat{n},0)}{r(\hat{q},\hat{n},\hat{n}) - r(\hat{q},\hat{n},0)}$. Notice that $\gamma \in [0,1]$. We then define the solution

$$\hat{\pi}(q,n,d) = \begin{cases} 0 & (q,n,d) = (\hat{q},\hat{n},\hat{d}) \\ \pi(\hat{q},\hat{n},\hat{n}) + \gamma\pi(\hat{q},\hat{n},\hat{d}) & (q,n,d) = (\hat{q},\hat{n},\hat{n}) \\ \pi(\hat{q},\hat{n},0) + (1-\gamma)\pi(\hat{q},\hat{n},\hat{d}) & (q,n,d) = (\hat{q},\hat{n},0) \\ \pi(q,n,d) & \text{else} \end{cases}$$

It follows that this solution satisfies (3.3d) and (3.3c). The change to the left-hand-side of (3.3a) is $\pi(\hat{q},\hat{n},\hat{d})(\gamma\hat{n}h(\hat{q},\hat{n},\hat{n}-1) - \hat{d}h(\hat{q},\hat{n},\hat{d}-1)) \geq 0$ by Lemma (3.4.2). The change in objective is

$$\gamma r(\hat{q},\hat{n},\hat{n}) + (1-\gamma)r(\hat{q},\hat{n},0) - r(\hat{q},\hat{n},\hat{d}) = \gamma(r(\hat{q},\hat{n},\hat{n}) - r(\hat{q},\hat{n},0)) - (r(\hat{q},\hat{n},\hat{d}) - r(\hat{q},\hat{n},0)) = 0,$$

and hence $\hat{\pi}$ has weakly higher utility. □

The proof of Theorem 3.2 depends on Lemma B.1, which states that the gain in firm utility from switching from telling n customers to join to telling m customers to join is greater in magnitude than any decrease in customer utility from the switch.

Lemma B.1. *Suppose conditioned on Q, the demands N_1 and N_2 are independent of each other. Then for any q and $n < m$ such that $h(q,n,n-1) < 0$ and $h(q,m,m-1) < 0$, we have*

$$\frac{mh(q,m,m-1)}{nh(q,n,n-1)} \leq \frac{r(q,m,m-1) - r(q,m,0)}{r(q,n,n-1) - r(q,n,0)} \tag{A2}$$

We now provide the proofs of Theorem 3.2 and Lemma B.1.

Proof of Theorem 3.2. By Theorem 3.1, the optimal signaling mechanism is a solution to the following linear program:

$$\max_{\pi(q,n,n),\pi(q,n,0)} \sum_{q,n} r(q,n,n)\pi(q,n,n) \tag{A3a}$$

$$\text{s.t.} \quad \sum_{q,n} nh(q,n,n-1)\pi(q,n,n) \geq 0 \tag{A3b}$$

$$\pi(q,n,n) + \pi(q,n,0) = \Phi(q,n) \quad \text{for all } q,n \tag{A3c}$$

$$\pi(q,n,0) = 0 \quad \text{for all } (q,n) \quad s.t. \quad h(q,n,n-1) \geq 0 \tag{A3d}$$

$$\pi(q,n,n), \quad \pi(q,n,0) \geq 0 \quad \text{for all } q,n \tag{A3e}$$

Consider any solution π to (A3). Suppose $\pi(\hat{q},\hat{n},\hat{n}) > 0$ and $\pi(\hat{q},\hat{m},0) > 0$ for $h(\hat{q},\hat{n},\hat{n}-1) < 0$, $h(\hat{q},\hat{m},\hat{m}-1) < 0$ and $\hat{n} < \hat{m}$. We can construct a solution $\hat{\pi}$ with higher objective. We define

$$\hat{\pi}(q,n,n) = \begin{cases} \pi(\hat{q},\hat{n},\hat{n}) - \rho \min\left(\frac{1}{\rho}\pi(\hat{q},\hat{n},\hat{n}), \pi(\hat{q},\hat{m},0)\right) & (q,n) = (\hat{q},\hat{n}) \\ \pi(\hat{q},\hat{m},m) + \min\left(\frac{1}{\rho}\pi(\hat{q},\hat{n},\hat{n}), \pi(\hat{q},\hat{m},0)\right) & (q,n) = (\hat{q},\hat{m}) \\ \pi(q,n,n) & \text{else} \end{cases}$$

and

$$\hat{\pi}(q,n,0) = \begin{cases} \pi(\hat{q},\hat{n},0) + \rho \min\left(\frac{1}{\rho}\pi(\hat{q},\hat{n},\hat{n}), \pi(\hat{q},\hat{m},0)\right) & (q,n) = (\hat{q},\hat{n}) \\ \pi(\hat{q},\hat{m},0) - \min\left(\frac{1}{\rho}\pi(\hat{q},\hat{n},\hat{n}), \pi(\hat{q},\hat{m},0)\right) & (q,n) = (\hat{q},\hat{m}) \\ \pi(q,n,n) & \text{else} \end{cases}$$

for

$$\rho = \frac{\hat{n}h(\hat{q},\hat{n},\hat{n}-1)}{\hat{m}h(\hat{q},\hat{m},\hat{m}-1)} \leq \frac{r(\hat{q},\hat{n},\hat{n}) - r(\hat{q},\hat{n},0)}{r(\hat{q},\hat{m},\hat{m}) - r(\hat{q},\hat{m},0)},$$

by Lemma B.1. We see that by construction, $\hat{\pi}$ satisfies (A3e) and (A3c). The change to the left-hand-side of (A3b) is

$$\min\left(\frac{1}{\rho}\pi(\hat{q},\hat{n},\hat{n}), \pi(\hat{q},\hat{m},0)\right)(\hat{m}h(\hat{q},\hat{m},\hat{m}-1) - \rho\hat{n}h(\hat{q},\hat{n},\hat{n}-1)) = 0,$$

so $\hat{\pi}$ satisfies (A3b) and is a feasible solution for (A3). The change in utility is

$$\min\left(\frac{1}{\rho}\pi(\hat{q},\hat{n},\hat{n}), \pi(\hat{q},\hat{m},0)\right)(r(\hat{q},\hat{m},\hat{m}) - r(\hat{q},\hat{m},0) - \rho(r(\hat{q},\hat{n},\hat{n}) - r(\hat{q},\hat{n},0))) \geq 0,$$

and thus $\hat{\pi}$ is a weakly better solution. Note that either $\hat{\pi}(\hat{q},\hat{n},\hat{n}) = 0$ or $\hat{\pi}(\hat{q},\hat{m},0) = 0$. We conclude that for any optimal π, if $\pi(\hat{q},\hat{n},\hat{n}) > 0$, then $\pi(\hat{q},\hat{m},0) = 0$ for all $m > n$ where $h(\hat{q},\hat{m},\hat{m}-1) < 0$ and $h(\hat{q},\hat{n},\hat{n}-1)$. Recalling that $\pi(q,n,d) = \sigma(q,n,d)\Phi(q,n)$, this implies that for any q, the optimal signaling mechanism σ has a threshold structure. □

Proof of Lemma B.1. From the proofs above, if $h(q,n,n-1) < 0$, then $q \geq n$. Thus, $\frac{r(q,m,m-1)-r(q,m,0)}{r(q,n,n-1)-r(q,n,0)} = \frac{m}{n}$. Next

$$\begin{aligned} h(q,m,m-1) &= (v-p_1) + c - \mathbf{E}\left[\min\left\{\frac{(Q-N_1+1)}{N_2+1}, 1\right\}(v-p_2)\middle| Q = q, N_1 = m\right] \\ &= (v-p_1) + c - \mathbf{E}\left[\min\left\{\frac{(Q-m+1)}{N_2+1}, 1\right\}(v-p_2)\middle| Q = q\right], \end{aligned}$$

where the second equality follows from the fact that N_1 and N_2 are independent conditioned on Q. By expanding and recalling that $h(q,n,n-1) < 0$, we see that (A2) is equivalent to

$$h(q,m,m-1) - h(q,n,n-1) \geq 0.$$

We see that, letting $f_2(n) = \mathbf{P}(N_2 = n|Q = q)$,

$$\begin{aligned}
&h(q,m,m-1) - h(q,n,n-1) = \\
&= -(v-p_2)\,\mathbf{E}\left[\min\left\{\frac{(Q-m+1)}{N_2+1},1\right\} - \min\left\{\frac{(Q-n+1)}{N_2+1},1\right\}\middle| Q=q\right] \\
&= -(v-p_2)\sum_{n_2=0}^{\infty} f_2(n_2)\left(\min\left\{\frac{(q-m+1)}{n_2+1},1\right\} - \min\left\{\frac{(q-n+1)}{n_2+1},1\right\}\right) \\
&= -(v-p_2)\left(\sum_{n_2=q-m+1}^{q-n} f_2(n_2)\left(\frac{q-m+1}{n_2+1}-1\right) + \sum_{n_2=q-n+1}^{\infty} f_2(n_2)\frac{-m+n}{n_2+1}\right) \\
&= (v-p_2)\left(\sum_{n_2=q-m+1}^{q-n} f_2(n_2)\left(1-\frac{q-m+1}{n_2+1}\right) + \sum_{n_2=q-n+1}^{\infty} f_2(n_2)\frac{m-n}{n_2+1}\right) \geq 0,
\end{aligned}$$

as desired. □

B.2 Construction of linear program

From Lemmas 3.1 and 3.2, the seller's decision problem (3.2) simplifies to finding a symmetric $\sigma : \Theta \times \{0,1\}^{\infty} \to [0,1]$ that maximizes the expected revenue subject to the requirement that obedience is an equilibrium:

$$\begin{aligned}
\max_{\sigma\in\Sigma} \quad & \mathbf{E}^{\sigma}\left[r(Q,N,D)\right] \\
\text{subject to,} \quad & \mathbf{E}^{\sigma}\left[Nh(Q,N,\widehat{D})\mathbf{I}\{s=1\}\right] \geq 0, \\
& \mathbf{E}^{\sigma}\left[Nh(Q,N,\widehat{D})\mathbf{I}\{s=0\}\right] \leq 0.
\end{aligned} \tag{A4}$$

Note that for any N and D, the probability a particular customer is asked to buy now is given by D/N, i.e., $\mathbf{P}(s=1|Q,N,D) = D/N$. Using this, it follows that

$$\begin{aligned}
\mathbf{E}^{\sigma}\left[Nh(Q,N,\widehat{D})\mathbf{I}\{s=1\}\right] &= \mathbf{E}^{\sigma}\left[Nh(Q,N,D-1)\mathbf{I}\{s=1\}\right] \\
&= \mathbf{E}^{\sigma}\left[Nh(Q,N,D-1)\,\mathbf{P}\left(s=1|Q,N,D\right)\right] \\
&= \mathbf{E}^{\sigma}\left[Nh(Q,N,D-1)\frac{D}{N}\right] \\
&= \sum_{q,n}\sum_{d=0}^{n}\Phi(q,n)\sigma(q,n,d)h(q,n,d-1)d.
\end{aligned} \tag{A5}$$

Here, in the first equality, we have used the fact that on the event $s = 1$, the number of other customers asked to buy now is given by $\widehat{D} = D - 1$. Similarly, using the fact that $\widehat{D} = D$ on the event that $s = 0$, we have

$$\mathbf{E}^{\sigma}\left[Nh(Q, N, \widehat{D})\mathbf{I}\{s = 0\}\right] = \sum_{q,n}\sum_{d=0}^{n} \Phi(q, n)\sigma(q, n, d)h(q, n, d)\,(n - d)\,. \tag{A6}$$

Finally, note that the objective of (A4) can be written as

$$\mathbf{E}\left[r(Q, N, D)\right] = \sum_{q,n}\sum_{d=0}^{n} r(q, n, d)\Phi(q, n)\sigma(q, n, d). \tag{A7}$$

Using (A5), (A6) and (A7), and after making a variable substitution where $\pi(q, n, d) \triangleq \Phi(q, n)\sigma(q, n, d)$, we can write (A4) as,

$$\max_{\pi} \sum_{q,n,d} r(q, n, d)\pi(q, n, d)$$

$$\text{subject to,} \sum_{q,n}\sum_{d=1}^{n} dh(q, n, d-1)\pi(q, n, d) \geq 0 \tag{A8a}$$

$$\sum_{q,n}\sum_{d=0}^{n-1} (n - d)\,h(q, n, d)\pi(q, n, d) \leq 0 \tag{A8b}$$

$$\sum_{d} \pi(q, n, d) = \Phi(q, n) \tag{A8c}$$

$$\pi(q, n, d) \geq 0, \quad \text{for all } q, n, d \tag{A8d}$$

B.3 Example of public signaling suboptimality with homogeneous customers

Consider a two-customer setting with equally likely states $Q \in \{0, 1\}$. Suppose customer utility is given by Figure B.1, and revenue is given by $\mathbf{I}(Q = D)$ when the state is Q and there are D customers buying. The optimal private signaling mechanism always instructs Q customers to buy (and chooses each with probability $\frac{1}{2}$ when $Q = 1$) using the signals "buy" and "wait" and achieves expected revenue 1. Note that it is incentive compatible for a customer to follow their signal: if a customer is told to buy, they know $Q = 1$ and thus want to buy. If a customer is told to wait, the probability of $Q = 0$ is

$$P(Q = 0|\text{wait}) = \frac{P(Q = 0, \text{wait})}{P(\text{wait})} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3},$$

and thus the optimal action is to wait.

We now calculate the optimal public signaling mechanism. Let σ be the revenue optimal public signaling mechanism with set of signals S, and a specific signal $s \in S$ such that the customer's posterior is given by $q = P(Q = 1|s)$. Suppose it is incentive compatible for a customer to buy with probability $p \in (0, 1)$. We see buying leads to expected utility $(1 - q)(-1) + q(1)$ and not buying leads to utility 0. Hence, $q = \frac{1}{2}$, and when $q = \frac{1}{2}$, any $p \in [0, 1]$ is an equilibrium.

We see the firm's revenue, given p, is

$$u(p) = (1 - q)(1 - p)^2 + q2p(1 - p).$$

We see

$$u'(p) = 2\left(-(1 - q) + q + p((1 - q) - 2q)\right).$$

Since $q = \frac{1}{2}$, $u'(0) = 0$ and $u'(1) = -2q < 0$. Since $u'(p)$ is a linear function, this implies $u'(p) < 0$ for all $p \in (0, 1]$. Hence, the revenue optimal choice of p given signal a is $p = 0$. This contradicts this being the revenue optimal public signaling mechanism.

Hence, in the revenue optimal public signaling mechanism, customers will always play a pure strategy. The revenue of such a mechanism can be no more than $\frac{1}{2}$, since when $Q = 1$, either all or no customers will buy, leading to revenue 0. The firm can then achieve the optimal public revenue by always instructing the customers to wait.

Thus, in this problem, private signaling achieves twice the revenue of any public signaling mechanism, even though the customers are homogeneous.

	Buy	Wait
Buy	$(-1,-1)$	$(-1,0)$
Wait	$(0,-1)$	$(0,0)$

(a) $Q=0$.

	Buy	Wait
Buy	$(1,1)$	$(1,0)$
Wait	$(0,1)$	$(0,0)$

(b) $Q=1$.

Figure B.1: Private signaling example.

B.4 Extensions to public signaling

In this section, we consider two extensions introduced in Section 3.6. First, we prove Theorem 3.3, which states that Theorem 3.1 holds when $N_2 = 0$ and the time 2 price is a function of remaining supply satisfying certain conditions. Next, we prove Theorem 3.4, which statesTheorem 3.1 holds when the initial inventory, Q, is a decision variable of the firm.

Proof of Theorem 3.3. To show this, we will show that Lemmas 3.3 and 3.4 apply. Then, the proof of Theorem 3.1 follows.

First, we show Lemma 3.3 applies in this setting: if $q < n$, we see that $h(q,n,n-1) = (v-p_1)\frac{q}{n} + c > 0$. If instead $q \geq n$, then

$$\begin{aligned} h(q,n,d-1) &= (v-p_1) + c - \min\left\{\frac{(q-d+1)}{n-d+1}, 1\right\}(v - p_2(q-d+1)) \\ &\leq (v-p_1) + c - \min\left\{\frac{(q-d)}{n-d}, 1\right\}(v - p_2(q-d)) = h(q,n,d) \end{aligned}$$

and thus $h(q,n,d-1) \geq 0$ implies $h(q,n,n-1) \geq 0$, as desired.

Finally, we show Lemma 3.4 applies: We first assume $h(q,n,n-1) \geq 0$. If $h(q,n,d-1) \leq 0$, then Lemma (3.4.1) holds. If not, suppose $h(q,n,d-1) \geq 0$. We would like to show that $dh(q,n,d-1) \leq (d+1)h(q,n,d)$ holds for all d where $h(q,n,d-1) \geq 0$. This would imply that $dh(q,n,d-1) \leq nh(q,n,n-1)$, as desired.

Now, suppose $q > d$. Then,

$$\begin{aligned} h(q,n,d-1) &= (v-p_1)+c-\min\left\{\frac{(q-d+1)}{n-d+1},1\right\}(v-p_2(q-d+1)) \\ &\le (v-p_1)+c-\min\left\{\frac{(q-d)}{n-d},1\right\}(v-p_2(q-d)) = h(q,n,d) \end{aligned}$$

Hence, $dh(q,n,d-1) \le (d+1)h(q,n,d)$. If $d \ge q$, then

$$dh(q,n,d-1) = d\left((v-p_1)\frac{q}{d}+c\right) \le (d+1)\left((v-p_1)\frac{q}{d+1}+c\right) = (d+1)h(q,n,d),$$

Thus $dh(q,n,d-1) \le (d+1)h(q,n,d)$ and $dh(q,n,d-1) \le nh(q,n,n-1)$, as desired.

Now suppose $h(q,n,n-1) < 0$. If $q < n$, we see that $h(q,n,n-1) = (v-p_1)\frac{q}{n}+c > 0$, which is a contradiction. Thus, $q \ge n$, and $h(q,n,d-1) = p_2(q-d+1)-p_1 \le 0$. We require

$$\frac{d}{n}\cdot h(q,n,d-1) \le \left(\frac{r(q,n,d)-r(q,n,0)}{r(q,n,n)-r(q,n,0)}\right)\cdot h(q,n,n-1).$$

Since $r(q,n,d)-r(q,n,0) = p_1 d + p_2(q-d)(n-d) - p_2(q)n$, we can rewrite the above condition as

$$n(p_2(q-n+1)-p_1)(p_1 d+p_2(q-d)(n-d)-p_2(q)n) \ge d(p_2(q-d+1)-p_1)(p_1 n-p_2(q)n).$$

This can be further simplified to

$$\begin{aligned} 0 \le p_1((n-d)(p_2(q)-p_2(q-d))+d(p_2(q-n+1)-p_2(q-d+1))) \\ + p_2(q)(dp_2(q-d+1)-np_2(q-n+1))+p_2(q-d)p_2(q-n+1)(n-d). \end{aligned}$$

Let A be the coefficient in front of p_1 and B be the remaining terms in the above. We will now show that both A and B are non-negative. Since $p_2(\cdot)$ is a convex function, $q-d = \frac{n-d-1}{n-1}q + \frac{d}{n-1}(q-n+1)$, and $q-d+1 = \frac{n-d}{n-1}q+\frac{d-1}{n-1}(q-n+1)$, we have

$$\begin{aligned} p_2(q-d) &\le \frac{n-d-1}{n-1}p_2(q)+\frac{d}{n-1}p_2(q-n+1), \\ p_2(q-d+1) &\le \frac{n-d}{n-1}p_2(q)+\frac{d-1}{n-1}p_2(q-n+1). \end{aligned}$$

Thus, with replacing $p_2(q-d)$ and $p_2(q-d+1)$, we have

$$\begin{aligned}
A &= (n-d)(p_2(q) - p_2(q-d)) + d(p_2(q-n+1) - p_2(q-d+1)) \\
&\geq \frac{1}{n-1}\left(p_2(q)(n-d)\left[(n-1)-(n-d+1)-d\right] + p_2(q-n+1)d\left[(n-1)-(d-1)-(n-d)\right]\right) \\
&\geq 0.
\end{aligned}$$

We see that

$$\begin{aligned}
B &= p_2(q)p_2(q-n+1)\left(d\frac{p_2(q-d+1)}{p_2(q-n+1)} + (n-d)\frac{p_2(q-d)}{p_2(q)} - n\right) \\
&= p_2(q)p_2(q-n+1)d(n-d)\left(\frac{\frac{1}{n-d}(p_2(q-d+1)-p_2(q-n+1))}{p_2(q-n+1)} + \frac{\frac{1}{d}(p_2(q-d)-p_2(q))}{p_2(q)}\right).
\end{aligned}$$

Since p_2 is convex, we see $\frac{1}{n-d}(p_2(q-d+1) - p_2(q-n+1)) \geq \Delta p_2(q-n+1)$ and $\frac{1}{d}(p_2(q)-p_2(q-d)) \leq \Delta p_2(q)$ (and thus $\frac{1}{d}(p_2(q-d)-p_2(q)) = -\frac{1}{d}(p_2(q)-p_2(q-d)) \geq -\Delta p_2(q)$). Using these, we get

$$\begin{aligned}
\frac{B}{p_2(q)p_2(q-n+1)d(n-d)} &= \frac{\frac{1}{n-d}(p_2(q-d+1)-p_2(q-n+1))}{p_2(q-n+1)} + \frac{\frac{1}{d}(p_2(q-d)-p_2(q))}{p_2(q)} \\
&\geq \frac{\Delta p_2(q-n+1)}{p_2(q-n+1)} - \frac{\Delta p_2(q)}{p_2(q)} \geq 0.
\end{aligned}$$

Thus

$$dh(q,n,d-1) \leq \left(\frac{r(q,n,d)-r(q,n,0)}{r(q,n,n)-r(q,n,0)}\right) nh(q,n,n-1)$$

and hence Lemma (3.4.2) holds. □

Proof of Theorem 3.4. We begin by proving Theorem (3.4.1), where the firm chooses a distribution, ρ, over Q before observing N. Let $\phi(\cdot)$ denote the distribution of N. Given any choice of ρ, Theorem 3.1 holds for distribution $\pi(q,n) = \rho(q)\phi(n)$. Hence, for *any choice* of ρ, the optimal signaling mechanism is public.

The proof of Theorem (3.4.2) is similar. Let $\rho(q|n)$ denote the probability of having q items when there are n customers in the first period. Once more, given any choice of ρ, Theorem 3.1 holds for distribution $\pi(q, n) = \phi(n)\rho(q|n)$. Hence, for any choice of ρ, the optimal signaling mechanism is public. $\square$

APPENDIX C

APPENDIX TO CHAPTER 4

C.1 Omitted Proofs

C.1.1 Proof of Lemma 4.3

To prove Lemma 4.3, we reference Lemma C.1, which proves the property for a simpler case.

Lemma C.1. *For any* $\hat{\omega} \in K_0$, $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\hat{\omega}\}) \subset \mathsf{Conv}(K_1 \cup K_{01}) \cup \mathsf{Conv}(K_{01} \cup \{\hat{\omega}\})$.

Proof. Let $\hat{\omega} \in K_0$ and consider any $\mu \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\hat{\omega}\})$. Suppose any decomposition of μ to members of $K_1 \cup K_{01} \cup \{\hat{\omega}\}$ necessarily assigns some weight to $\hat{\omega}$ and some weight to members of K_1. Consider the decomposition that assigns weight to the fewest members of K_1. Letting $L \subset K_1$ denote the set of elements assigned positive weight, we can then write

$$\mu = \alpha_{\hat{\omega}}\hat{\omega} + \sum_{\omega \in L} \alpha_\omega \omega + \sum_{\phi \in K_{01}} \alpha_\phi \phi.$$

Let $\omega' \in L$ be chosen arbitrarily, and let $\psi = \frac{\alpha_{\hat{\omega}}\hat{\omega} + \alpha_{\omega'}\omega'}{\alpha_{\hat{\omega}} + \alpha_{\omega'}} \in \Delta(\Omega)$. Recall that $\chi(\hat{\omega}, \omega')$ is the distribution over $\hat{\omega}$ and ω' in K_{01}. Since ψ is a convex combination of $\hat{\omega}$ and ω', it is also either a convex combination of (1) ω' and $\chi(\hat{\omega}, \omega')$ or (2) $\chi(\hat{\omega}, \omega')$ and $\hat{\omega}$. In the first case, $\psi \in \mathsf{Conv}(K_1 \cup K_{01})$ and we can write

$$\mu = (\alpha_{\hat{\omega}} + \alpha_{\omega'})\psi + \sum_{\substack{\omega \in L \\ \omega \neq \omega'}} \alpha_\omega \omega + \sum_{\phi \in K_{01}} \alpha_\phi \phi \in \mathsf{Conv}(K_1 \cup K_{01}).$$

In the second case, we can write $\psi = \gamma\hat{\omega} + (1-\gamma)\phi'$, for some $\gamma \in [0, 1]$. Then, we can express μ as a convex combination over fewer members of K_1 (since we have removed the weight on ω') which is a contradiction. Thus, $\mu \in \mathsf{Conv}(K_1 \cup K_{01}) \cup \mathsf{Conv}(K_{01} \cup \{\hat{\omega}\})$ and $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\hat{\omega}\}) = \mathsf{Conv}(K_1 \cup K_{01}) \cup \mathsf{Conv}(K_{01} \cup \{\hat{\omega}\})$. $\square$

With Lemma C.1, we can prove Lemma 4.3.

Proof of Lemma 4.3. Consider any $\mu \in \Delta(\Omega)$. We can write μ as a convex combination of the elements of $K_0 \cup K_1 = \Omega$. Thus, we can also write μ as a convex combination of $K_0 \cup K_1 \cup K_{01}$. Similar to before, let this be the decomposition that places positive weight on the fewest number of elements of K_0:

$$\mu = \sum_{\omega \in K_0} \alpha_\omega \omega + \sum_{\omega \in K_1 \cup K_{01}} \alpha_\omega \omega.$$

Let $\hat{\omega}$ denote any element of K_0 such that $\alpha_\omega > 0$. We see we can write

$$\begin{aligned}
\mu &= \sum_{\substack{\omega \in K_0 \\ \omega \neq \hat{\omega}}} \alpha_\omega \omega + \alpha_{\hat{\omega}} \hat{\omega} + \sum_{\omega \in K_1 \cup K_{01}} \alpha_\omega \omega \\
&= \sum_{\substack{\omega \in K_0 \\ \omega \neq \hat{\omega}}} \alpha_\omega \omega + A \left(\frac{\alpha_{\hat{\omega}}}{A} \hat{\omega} + \sum_{\omega \in K_1 \cup K_{01}} \frac{\alpha_\omega}{A} \omega \right)
\end{aligned}$$

where $A = \left(\sum_{\omega \in K_1 \cup K_{01} \cup \{\hat{\omega}\}} \alpha_\omega \right)$. Let $\psi = \frac{\alpha_{\hat{\omega}}}{A} \hat{\omega} + \sum_{\omega \in K_1 \cup K_{01}} \frac{\alpha_\omega}{A} \omega$. ψ is an element of $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\hat{\omega}\})$, and by Lemma C.1, $\psi \in \mathsf{Conv}(K_1 \cup K_{01}) \cup \mathsf{Conv}(K_{01} \cup \{\hat{\omega}\})$. If $\psi \in \mathsf{Conv}(K_{01} \cup \{\hat{\omega}\})$, then we have expressed μ as an element of $\mathsf{Conv}(K_0 \cup K_{01})$, as desired. If not, $\psi \in \mathsf{Conv}(K_1 \cup K_{01})$ and we can express μ while putting positive weight on fewer elements of K_0. This is a contradiction. Thus, $\mu \in \mathsf{Conv}(K_0 \cup K_{10}) \cup \mathsf{Conv}(K_1 \cup K_{10})$, and we conclude $\Delta(\Omega) = \mathsf{Conv}(K_0 \cup K_{01}) \cup \mathsf{Conv}(K_1 \cup K_{01})$. $\square$

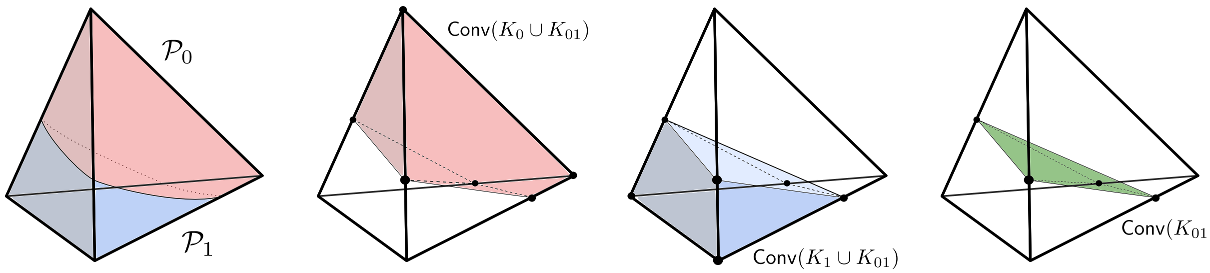


Figure C.1: Geometry of $\Delta(\Omega)$ when $\mathsf{Conv}(K_{01})$ is not a hyperplane.

C.1.2 Additional Figures

C.1.3 Queueing analogue of theorem 4.3

We can write the choice of an optimal signaling scheme as the convex program

$$\begin{aligned}\max_{t_0,t_1} \ & \sum_{\omega\in\Omega} t_1(\omega) \\ \text{subject to,} \quad & t_0 \in \mathsf{Conv}(\widehat{\mathcal{P}}_0), \\ & t_1 \in \mathsf{Conv}(\widehat{\mathcal{P}}_1), \\ t_0(\omega+1)+t_1(\omega+1) &= \lambda t_1(\omega), \qquad \text{for each } \omega\in\Omega.\end{aligned} \tag{A1}$$

We can, like in Section 4.4, write this convex program as a linear program. In an analogue to Theorem 4.3, we have the following theorem.

Theorem C.1. *When $\bar{\rho}$ satisfies Assumption 4.2 and K_1 is finite, the sender's persuasion problem can be optimized by solving the following linear program:*

$$\max_{t_0,t_1} \sum_{\omega\in\Omega} t_1(\omega)$$

$$\textit{subject to,} \quad t_0 \in \mathsf{Conv}(K_0 \cup K_{01} \cup \{\mathbf{0}\}), \tag{A2a}$$

$$t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\}), \tag{A2b}$$

$$t_0(\omega+1)+t_1(\omega+1) = \lambda t_1(\omega), \qquad \textit{for each } \omega\in\Omega. \tag{A2c}$$

To prove this theorem, we need a result analogous to Lemma 4.3,

Lemma C.2. *In a queue with finite K_1, $\Delta(\Omega) = \mathsf{Conv}(K_0 \cup K_{01}) \cup \mathsf{Conv}(K_1 \cup K_{01})$.*

Proof of Lemma C.2. Note Lemma C.1 applies to our setting from the finiteness of K_1: for any $\hat{\omega} \in K_0$, $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\hat{\omega}\}) \subset \mathsf{Conv}(K_1 \cup K_{01}) \cup \mathsf{Conv}(K_{01} \cup \{\hat{\omega}\})$.

Consider any $\mu^* = \sum_{\omega \in K_0} \alpha_\omega \omega + \sum_{\omega \in K_1} \alpha_\omega \omega \in \Delta(\Omega)$. Let $\mu_0^0 = \sum_{\omega \in K_0} \alpha_\omega \omega$ and $\mu_0^1 = \sum_{\omega \in K_1} \alpha_\omega \omega \in \Delta(\Omega)$. Note that $\mu_0^0 \in \mathsf{Conv}(K_0 \cup \{\mathbf{0}\})$, $\mu_0^1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$.

We now define sequences $\{\mu_n^0\}$ and $\{\mu_n^1\}$ with the properties: $\mu^* = \mu_n^0 + \mu_n^1$ and $\mu_n^0 = \sum_{\omega \in K_1} \beta_\omega \omega \mathsf{Conv}(K_0 \cup \{\mathbf{0}\})$.We choose the smallest $\hat{\omega} \in K_1$ such that $\beta_\omega > 0$; if no such $\hat{\omega}$ exists, we let $\mu_{n+1}^0 = \mu_n^0 = 0$ and $\mu_{n+1}^1 = \mu_n^1$. We then define

$$\mu_{n+1}^0 = \mu_n^0 - \beta_{\hat{\omega}} \hat{\omega}$$
$$\mu_{n+1}^1 = \mu_n^1 + \beta_{\hat{\omega}} \hat{\omega}.$$

Note that since these two properties hold for μ_0^0 and μ_0^1, they hold for all members of this sequence. Next, suppose that $\mu_n^1 = \gamma\mu \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, where $\mu \in \mathsf{Conv}(K_1 \cup K_{01})$. We see that, by Lemma C.1,

$$\frac{\mu_{n+1}^1}{\gamma + \beta_{\hat{\omega}}} = \frac{\mu_n^1 + \beta_{\hat{\omega}} \hat{\omega}}{\gamma + \beta_{\hat{\omega}}} \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\hat{\omega}\}) \subset \mathsf{Conv}(K_1 \cup K_{01}) \cup \mathsf{Conv}(K_{01} \cup \{\hat{\omega}\}).$$

If $\frac{\mu_{n+1}^1}{\gamma + \beta_{\hat{\omega}}} \in \mathsf{Conv}(K_{01} \cup \{\hat{\omega}\})$, then $\mu_{n+1}^1 \in \mathsf{Conv}(K_0 \cup K_{01} \cup \{\mathbf{0}\})$ and $\mu^* = \mu_{n+1}^1 + \mu_{n+1}^0 \in \mathsf{Conv}(K_0 \cup K_{01})$, as desired. If not, then $\frac{\mu_{n+1}^1}{\gamma + \beta_{\hat{\omega}}} \in \mathsf{Conv}(K_1 \cup K_{01})$, and we can repeat this process, again with the property that $\mu_{n+1}^1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$.

If ever this sequence has $\mu_{n+1}^1 \in \mathsf{Conv}(K_0 \cup K_{01} \cup \{\mathbf{0}\})$, then we have $\mu^* = \mu_{n+1}^1 + \mu_{n+1}^0 \in \mathsf{Conv}(K_0 \cup K_{01})$, as desired. The only case that remains is if this does not occur for any value of n.

In this case, $\mu_n^1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$ for all values of n. Since $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$ is a closed set, its limit point is in $\mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$ as well. The only such limit point is μ^*, which means $\mu^* \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, and the lemma is proven. □

With this, we can prove the Theorem.

Proof of Theorem C.1. The proof is identical to that of Theorem 4.3, but with Lemma C.2 referenced instead of Lemma 4.3. □

C.1.4 Proof of Theorem 4.4

We prove this theorem in two steps, like the comparable proof in Lingenbrink and Iyer (2018a). First, we show that the optimal signaling scheme would signal a receiver to join the queue if she would have joined under full-information. With this structure in place, we then show that any feasible solution that does not have a threshold structure can be perturbed to obtain another feasible solution corresponding to a threshold scheme with equal or higher throughput.

Let $M = \max(K_1)$ be the largest queue length where the receiver would join under full information. Consider any feasible solution, t, to (A2), where $t_0(n') > 0$ for some $n' \leq M$. We will construct a feasible solution, $\hat{t}$ to (A2) that has higher revenue than t and satisfies $\hat{t}_0(n) = 0$ for all $n \leq M$. We define

$$\hat{t}_1(n) = \begin{cases} \frac{1}{Z}(t_0(0) + t_1(0))\lambda^n & \text{for } n \leq M \\ \frac{1}{Z} t_1(n) & \text{for } n > M \end{cases}$$

,

$$\hat{t}_0(n) = \begin{cases} 0 & \text{for } n \leq M \\ \frac{1}{Z}\left((t_0(0) + t_1(0))\lambda^{M+1} - t_1(M+1)\right) & \text{for } n = M \\ \frac{1}{Z} t_0(n) & \text{for } n > M+1 \end{cases},$$

where $Z = (t_0(0) + t_1(0)) \sum_{n=0}^{M+1} \lambda^n + \sum_{n=M+1}^{k-1} t_0(n) + t_1(n)$ is a normalizing constant to ensure that $\hat{t}$ is a proper probability distribution. Since $(t_0(0) + t_1(0))\lambda^n \geq t_1(n) + t_0(n)$, we see

$Z \geq 1$. Since $t_0(n') > 0$, $(t_0(0) + t_1(0))\lambda^{n'+1} > \lambda t_1(n') = t_1(n'+1) + t_0(n'+1)$, and thus $Z > 1$.

We first show that $\hat{t}$ is a feasible solution. We start by ensuring (A2a) holds: Note that $t_1(k) \leq (t_0(0) + t_1(0))\lambda^n$ for all n, and hence $\hat{t}_1(n) \geq t_1(n)/Z$ for all n. More precisely, $\hat{t}_1(n) \geq t_1(n)/Z$ for $n \in K_1$ and $\hat{t}_1(n) = t_1(n)/Z$ for $n \notin K_1$. Let $\nu_1 = \frac{Z-1}{Z}\left(\hat{t}_1 - t_1/Z\right)$ denote the normalized difference between $\hat{t}_1$ and t_1/Z. We see $\nu_1 \in \mathsf{Conv}(K_1 \cup \{0\}) \subset \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, and hence $\hat{t}_1 = \frac{1}{Z}t_1 + \frac{Z-1}{Z}\nu_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, as desired.

Next, notice that $\hat{t}_0 \in \mathsf{Conv}(K_0 \cup \{0\}) \subset \mathsf{Conv}(K_0 \cup K_{01} \cup \{\mathbf{0}\})$, by definition. Hence, (A2b) is satisfied.

We see that (A2c) is satisfied: for $n < M$,

$$\lambda\hat{t}_1(n) = \frac{1}{Z}(t_0(0) + t_1(0))\lambda^{n+1} = \hat{t}_1(n+1) + 0 = \hat{t}_1(n) + \hat{t}_0(n);$$

for $n = M$,

$$\lambda\hat{t}_1(M) = \frac{1}{Z}(t_0(0)+t_1(0))\lambda^{M+1} = \frac{1}{Z}\left((t_0(0) + t_1(0))\lambda^{M+1} - t_1(n+1) + t_1(n+1)\right) = \hat{t}_1(M+1)+\hat{t}_0(M+1);$$

and for $n > M$,

$$\lambda\hat{t}_1(n) = \frac{1}{Z}(\lambda t_1(n)) = \frac{1}{Z}\left(t_1(n+1) + t_{0,n+1}\right) = \hat{t}_1(n+1) + \hat{t}_0(n+1).$$

Since $Z > 1$, we see the difference in utility between $\hat{t}$ and t is

$$1 - \hat{t}_0(0) - \hat{t}_1(t) - (1 - t_0(0) - t_1(1)) = \left(1 - \frac{1}{Z}\right)(t_0(0) + t_1(1)) > 0.$$

Henceforth, we restrict to feasible solutions t that satisfy the property $t_0(n) = 0$ for $n \leq M$. Note that any t that satisfies this condition, (A2a) and (A2c) naturally satisfies (A2b) because $t_0(n) = 0$ for $n \notin K_0$. Hence, $t_0 \in \mathsf{Conv}(K_0\{0\}) \subset \mathsf{Conv}(K_0 \cup K_{01} \cup \{\mathbf{0}\})$, and we can ignore the constraint (A2b) in our analysis.

Consider now a feasible solution t such that there exists an $N > M$ such that $t_0(n) = 0, t_1(n) = \lambda^n t_1(0)$ for $n < N$, $0 < t_1(N) < \lambda t_1(N-1)$, $t_0(N) > 0$ and $t_1(N+1) > 0$. We will construct a perturbation and show it remains feasible and achieves the same objective value.

Let $\hat{t}$ be defined as

$$\hat{t}_1(n) = \begin{cases} t_1(n) & \text{for } n < N \\ t_1(N) + \beta \sum_{i=N+1}^{\infty} t_1(i) & \text{for } n = N \\ (1-\beta) t_1(n) & \text{for } n > N \end{cases},$$

$$\hat{t}_0(n) = \begin{cases} t_0(n) & \text{for } n < N \\ t_0(N) - \beta \sum_{i=N+1}^{\infty} t_1(i) & \text{for } n = N \\ t_0(N+1) + \beta t_1(N+1) + \lambda\beta \sum_{i=N+1}^{\infty} t_1(i) & \text{for } n = N+1 \\ (1-\beta) t_0(n) & \text{for } n > N \end{cases},$$

for some $\beta \in (0, 1]$ to be defined later. This is chosen so it mimics the analysis in Chapter 2. This ensures no change in utility. We begin by showing that $\hat{t}_1 \in \mathsf{Conv}(\mathcal{P}_1 \cup \{0\})$.

Recall that each element of K_{01} can be written as $\chi(m, n) = \gamma(m, n)m + (1 - \gamma(m, n))n$ for $0 \le n \le M$ and $m > M$.

Since $t_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$, we can write the decomposition

$$t_1 = \sum_{n=0}^{M} \alpha_n n + \sum_{n=0}^{M} \sum_{m=M+1}^{\infty} \alpha_{m,n} \chi(m, n).$$

Consider

$$\tilde{t}_1 = \sum_{n=0}^{M} \tilde{\alpha}_n n + \sum_{n=0}^{M} \sum_{m=M+1}^{\infty} \tilde{\alpha}_{m,n} \chi(m, n),$$

where, for any $n \le M$,

$$\tilde{\alpha}_n = \alpha_n + \beta \sum_{m=N+1}^{\infty} \alpha_{m,n} \left(1 - \frac{\gamma(m, n)}{\gamma(N, n)}\right),$$

and

$$\tilde{\alpha}_{m,n} = \begin{cases} \alpha_{m,n} & \text{for } m < N \\ \alpha_{N,n} + \beta \sum_{\ell=N+1}^{\infty} \frac{\gamma(\ell,n)}{\gamma(N,n)} \alpha_{\ell,n} & \text{for } m = N \\ (1-\beta)\alpha_{m,n} & \text{for } m > N \end{cases} \cdot$$

Note that $\gamma(\ell, n) \leq \gamma(N, n)$ for all $\ell > N$ and for any $\beta \in (0, 1]$, these coefficients are non-negative. Next, notice that the sum of the coefficients is

$$\begin{aligned} S =& \sum_{n=0}^{M} \left(\tilde{\alpha}_n + \sum_{m=M+1}^{\infty} \tilde{\alpha}_{m,n} \right) \\ =& \sum_{n=0}^{M} \left(\alpha_n + \beta \sum_{m=N+1}^{\infty} \alpha_{m,n} \left(1 - \frac{\gamma(m,n)}{\gamma_{N,n}} \right) + \sum_{m=M+1}^{N-1} \alpha_{m,n} + \sum_{m=N+1}^{N-1} \beta \frac{\gamma(m,n)}{\gamma(N,n)} \alpha_{m,n} + \sum_{m=N+1}^{N-1} (1-\beta) \alpha_{m,n} \right) \\ \leq& 1. \end{aligned}$$

Hence, $\tilde{t}_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$. It remains to show that $\hat{t}_1 = \tilde{t}_1$. We see that, for $n \leq M$,

$$\begin{aligned} \tilde{t}_1(n) - \hat{t}_1(n) &= \hat{\alpha}_n + \sum_{m=M+1}^{\infty} (1 - \gamma(m,n)) \hat{\alpha}_{m,n} - \left(\alpha_n + \sum_{m=M+1}^{\infty} (1 - \gamma(m,n)) \alpha_{m,n} \right) \\ &= \beta \sum_{m=N+1}^{\infty} \alpha_{m,n} \left(1 - \frac{\gamma(m,n)}{\gamma(N,n)} \right) + \frac{1 - \gamma(N,n)}{\gamma(N,n)} \beta \sum_{m=N+1}^{\infty} \gamma(m,n) \alpha_{m,n} - \beta \sum_{m=N+1}^{\infty} (1 - \gamma(m,n)) \alpha_{m,n} \\ &= 0. \end{aligned}$$

Next, for $M + 1 \leq m < N$,

$$\tilde{t}_1(m) - \hat{t}_1(m) = \sum_{i=0}^{M} \gamma(m,n) \tilde{\alpha}_{m,n} - \sum_{i=0}^{M} \gamma(m,n) \alpha_{m,n} = 0.$$

For $m = N$, we see

$$\begin{aligned} \tilde{t}_1(N) - \hat{t}_1(N) &= \sum_{i=0}^{M} \gamma(N,n) \tilde{\alpha}_{N,n} - t_1(N) - \beta \sum_{m=N+1}^{\infty} t_1(m) \\ &= \beta \sum_{i=0}^{M} \gamma(N,n) \sum_{m=N+1}^{\infty} \frac{\gamma(m,n)}{\gamma(N,n)} \alpha_{m,n} - \beta \sum_{m=N+1}^{\infty} \sum_{n=0}^{M} \gamma(m,n) \alpha_{m,n} \\ &= 0. \end{aligned}$$

Finally, for $m > N$, we see

$$\tilde{t}_1(m) - \hat{t}_1(m) = (1 - \beta) \sum_{i=0}^{M} \gamma(m,n) \alpha_{m,n} - (1 - \beta) \sum_{i=0}^{M} \gamma(m,n) \alpha_{m,n} = 0.$$

Thus, $\hat{t}_1 = \tilde{t}_1 \in \mathsf{Conv}(K_1 \cup K_{01} \cup \{\mathbf{0}\})$. Further, assuming $\beta \leq \frac{t_0(N)}{\sum_{i=N+1}^{\infty} t_1(i)}$, $\hat{t}$ satisfies (A2c), by construction. Thus, we have constructed a solution with a equal objective value.

If $\frac{t_0(N)}{\sum_{i=N+1}^{\infty} t_1(i)} \geq 1$, then choosing $\beta = 1$ yields $t_1(n) = 0$ for all $n > N$. If $\frac{t_0(N)}{\sum_{i=N+1}^{\infty} t_1(i)} < 1$, then choosing $\beta = \frac{t_0(N)}{\sum_{i=N+1}^{\infty} t_1(i)} \geq 1$ yields $t_0(N) = 0$. We then obtain that any t such that there exists an $N > M$ such that $t_0(n) = 0, t_1(n) = \lambda^n t_1(0)$ for $n < N$, $0 < t_1(N) < \lambda t_1(N-1)$, $t_0(N) > 0$ and $t_1(N+1) > 0$ can be perturbed to obtain a solution $\hat{t}$ satisfying either (1) $t_0(n) = 0, t_1(n) = \lambda^n t_1(0)$ for $n < N$, $0 < t_1(N) \leq \lambda t_1(N-1)$ and $t_1(n) = 0$ for $n > N$ or (2) $t_0(n) = 0, t_1(n) = \lambda^n t_1(0)$ for $n \leq N$. By induction, this implies that if the optimum is attained, it is attained by a feasible solution for which there exists an $N \geq M$ such that $t_0(n) = 0, t_1(n) = \lambda^n t_1(0)$ for all $n < N$, $0 < t_1(N) \leq \lambda t_1(N-1)$, and $t_1(n) = 0$ for all $n > N$.

C.1.5 Proof of Lemma 4.5

Consider two distributions $\mu, \nu \in \Delta(\Omega)$. Let $\kappa = \gamma\mu + (1-\gamma)\nu$ for some $\gamma \in [0,1]$. Let $\hat{\mu}$ be a random variable that is μ with probability γ and ν with probability $(1-\gamma)$. Then, by the law of total variance,

$$\mathrm{Var}_\kappa[X] = \mathbf{E}_{\hat{\mu}}[\mathrm{Var}(X)] + \mathrm{Var}_{\hat{\mu}}(\mathbf{E}[X]) \geq \gamma \,\mathrm{Var}_\mu[X] + (1-\gamma)\,\mathrm{Var}_\nu[X],$$

and thus $\mathrm{Var}_\mu[X]$ is concave in μ. Further, $\sqrt{\cdot}$ is a concave, non-decreasing function, so their composition is concave. Since expectation is linear, $\bar{\rho}(\mu) = \tau - \left(\mathbf{E}_\mu[X] + \beta\sqrt{\mathrm{Var}_\mu[X]}\right)$ is convex.

C.1.6 Statement and Proof of Proposition C.1

Proposition C.1. *If $1+\beta > \tau$, then under any signaling mechanism, throughput is 0. If $1+\beta \leq \tau$, then let ω^* be the largest integer such that $\omega + 1 + \beta\sqrt{\omega+1} \leq \tau$. Under the fully-revealing mechanism, throughput is $1-(\lambda-1)/(\lambda^{\omega^*+2}-1)$ if $\lambda \neq 1$ and $1-1/(\omega^*+2)$ if $\lambda = 1$. Under the no-information mechanism, throughput is $\min\{\lambda, 1-\frac{1+\beta}{\tau}\}$.*

Proof. If $1+\beta > \tau$, then under any signaling scheme, no customer joins the queue because even if the queue is empty and the customer is serviced immediately, customer utility is $\tau - (1+\beta\sqrt{1}) < 0$. Therefore, throughput is 0. Henceforth, we assume that $1+\beta \leq \tau$. Then, $\omega^* \geq 0$ as defined in the proposition exists.

When the queue length is ω, X_μ is a sum of $\omega+1$ independent random variables, each exponentially distributed with rate 1 (the waiting times for ω people in the queue plus the customer's own service time). Therefore, $\mathbf{E}[X_\mu|\omega] = \omega+1$ and $\mathrm{Var}(X_\mu|\omega) = \omega+1$.

Under the fully-revealing scheme, the customer knows the queue length ω and joins if and only if $\tau - (\mathbf{E}[X_\mu|\omega] + \beta\,\mathrm{Var}(X_\mu|\omega)) = \tau - (\omega+1+\beta\sqrt{\omega+1}) \geq 0$. That is, the customer joins the queue if and only if $\omega \leq \omega^*$. Therefore, the queue is an $M/M/1$ queue with maximum queue length ω^*+1 such that customers arrive according to a Poisson process with rate λ and customers are turned away if the queue is full. Standard result in queuing theory gives the throughput $1-1/(1+\lambda+\cdots+\lambda^{\omega^*+1})$, which is $1-(\lambda-1)/(\lambda^{\omega^*+2}-1)$ if $\lambda \neq 1$ and $1-1/(\omega^*+2)$ if $\lambda = 1$.

Under the no-information scheme, the customer strategy is to join the queue with probability $q \in [0,1]$. $q = 0$ is an equilibrium if and only if $1+\beta > \tau$, which we rule out. We can view the queue as a thinned $M/M/1$ queue with arrival rate $q\lambda \in (0,1)$. Now we compute the customer utility for joining the queue. Standard results on the $M/M/1$ queue give the queue throughput $q\lambda$, and $\Pr(\omega) = (1-q\lambda)(q\lambda)^\omega$, so ω has a geometric distribution, and $\mathbf{E}[\omega] = \frac{q\lambda}{1-q\lambda}$ and $\mathrm{Var}(\omega) = \frac{q\lambda}{(1-q\lambda)^2}$. (The expectation and variance of ω can also be computed directly from the probability mass density of ω.) Therefore, by the law of iterated expectations and the law of total variance,

$$\mathbf{E}[X_\mu] = \mathbf{E}[\mathbf{E}[X_\mu|\omega]] = \mathbf{E}[\omega + 1] = \frac{1}{1-q\lambda}$$

$$\mathrm{Var}(X_\mu) = \mathbf{E}[\mathrm{Var}(X_\mu|\omega)] + \mathrm{Var}(\mathbf{E}[X_\mu|\omega]) = \mathbf{E}[\omega+1] + \mathrm{Var}(\omega+1) = \frac{1}{1-q\lambda} + \frac{q\lambda}{(1-q\lambda)^2} = \frac{1}{(1-q\lambda)^2}$$

Therefore, the customer utility for joining the queue is $\tau - (\mathbf{E}[X_\mu] + \beta\,\mathrm{Var}(X_\mu)) = \tau - \frac{1+\beta}{q\lambda}$.

$q = 1$ is an equilibrium if and only if the customer utility for joining the queue is weakly higher than the customer utility for not joining the queue, that is, $\tau - \frac{1+\beta}{\lambda} \geq 0$ or $\frac{1}{\lambda}(1 - \frac{1+\beta}{\tau}) \geq 1$. $q \in (0,1)$ is an equilibrium if and only if the customer utility for joining the queue is equal to the customer utility for not joining the queue, so $q = \frac{1}{\lambda}(1 - \frac{1+\beta}{\tau})$. Therefore, if $1 + \beta \leq \tau$, then the no-information equilibrium is $q = \min\{1, \frac{1}{\lambda}(1 - \frac{1+\beta}{\tau})\}$ and the queue throughput is $q\lambda = \min\{\lambda, 1 - \frac{1+\beta}{\tau}\}$.

□